Mechanisms of Transcriptional Control in Phosphate-responsive Signaling Pathway of Saccharomyces cerevisiae

Mechanisms of Transcriptional Control in Phosphate-responsive Signaling Pathway of *Saccharomyces cerevisiae*

A dissertation presented

By

Xu Zhou

To

The Department of Molecular and Cellular Biology

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

July 2012

Dissertation Advisor: Professor Erin K. O'Shea Xu Zhou

Mechanisms of Transcriptional Control in the Phosphate-responsive Signaling Pathway of *Saccharomyces cerevisiae*

Abstract

Regulation of gene expression is essential for many biological processes. Binding of transcription factors to DNA is a key regulatory step in the control of gene expression. It is commonly observed that DNA sequences with high affinity for transcription factors occur more frequently in the genome than the instances of genes bound or regulated by these factors. However, the mechanism by which transcription factors selectively identify and regulate these genes was unclear. I utilized the transcriptional control of the phosphate-responsive signaling pathway (PHO) in *Saccharomyces cerevisiae* as a model system to address this problem.

I applied genome-wide approaches to study how the intrinsic and extrinsic influences shape the binding and regulatory landscape of Pho4, a yeast transcription factor that activates gene expression in response to phosphate limitation. I demonstrated that the DNA binding affinity of Pho4 is necessary, but not sufficient, to dictate Pho4 binding *in vivo* - only a subset of its high affinity binding sites are bound. I found that nucleosomes significantly restrict the sites to which Pho4 binds. At nucleosome-depleted sites, competition between Pho4 and another transcription factor, Cbf1, determines Pho4 occupancy. Competition from Cbf1 also regulates transcription of PHO response: it raises the threshold for transcriptional activation by Pho4 in phosphate replete

conditions and prevents Pho4 from activating genes outside the phosphate regulon during phosphate starvation. Furthermore, Pho4 binding is not sufficient for transcriptional activation. By quantitatively dissecting the regulatory interaction between Pho4 and its cofactor Pho2, I discovered that genes activated by Pho4 require cooperative binding with Pho2 at their promoters. Combining these experimental observations, I was able to predict Pho4 binding and its functionality at the whole genome scale. Novel phosphate-responsive anti-sense transcripts were identified adjacent to the predicted functional binding sites that are not associated with a gene.

This work demonstrated that the specificity of the PHO pathway transcriptional control is combinatorially determined by the intrinsic DNA binding affinity of Pho4 and extrinsic factors that compete and cooperate with Pho4. This work provided insights into the mechanisms of global control by sequence-specific transcription factors and a roadmap to dissect eukaryotic transcriptional networks.

Table of Contents

To those who accompanied me for the past 6 years –

My friends, family

and

budding yeasts

Acknowledgement

The success of any scientific project depends on the encouragement and guidance of many others. Throughout the years, the number of people who have crossed path with my academic and extracurricular life can hardly be counted. Here is an incomplete list of my attempt to recognizing people who have been instrumental in the successful completion of my thesis.

First and foremost, I owe my deepest gratitude to my advisor Erin O'Shea, who has guided me with her diligence and rigorous scientific attitude throughout my graduate training. I have learned from her not only how to dissect and probe a scientific problem, but more importantly, how to explore and discover questions yet to be answered. I will forever miss her insightful critiques and the stimulating environment that she brought together, as I move on with my scientific career.

The present and the past members of the O'Shea lab have been an essential of my life both in and outside of academia. They are the true source of knowledge and humor, and caring friends to share my joy and frustration. I am grateful to have the opportunity to know each of them: Peter Arvidson, Andrew Capaldi, Joonhyuk Choi, Christopher Chidley, Kristen Cook, Elmar Czeko, Siting Gan, Catherine Grimes, Andrian Gutu, Anders Hansen 'the angry Norwegian', Nan Hao, Jeremy Johnson, Felex Lam, Young-Sam Lee, Harold Kim, Joe Markson, Shankar Mukerji, Tim Peterson, Joe Piehura, Aditya Paul, Anna Puszynska, Abbhi Rajagopal, Abbas Rizvi, Mike Rust, Shuwen Teng, Vikram Vijayan, Brian Zid, Rick Zuzow, Shari Grossman, Isha Jain, Shawn Low, Charlotte Seid, Tim Schmidt, Lynn Yi. Our neighbors, the members of Denic lab, have also brought me tremendous help and friendships, and occasionally birthday celebration cakes. I am indebted to the Bauer core facility and staffs for their technical

support. Particularly, I would like to thank Jiangwen Zhang and Christian Daly for support of high-throughput sequencing that was an essential part of my dissertation.

I would like to express the deepest appreciation to my dissertation committee members, Professor Andrew Murray, Professor Aviv Regev and Professor Vlad Denic. Particularly, Andrew transformed my anxiety about the qualification exam into excitement of defending intellectual challenges, and guided me with his humor and sparks of ideas. Advice given by Aviv has been a great help in both my graduate work and my exploration of further interests in academia. Vlad has been the go-to person of my experimental problems and the master who led me into the world of caffination. In addition, I would like to thank Professor Richard Losick for being on my qualification exam committee and Professor Susan Mango for being on my third year dissertation committee. Without the guidance and persistent help from all of them, this dissertation would not have been possible.

I am also very grateful to my classmates and friends from the department of Molecular and Cellular Biology and the administrative staffs for cultivating a nurturing environment so that I would not feel lonely thousands of miles away from home.

Last but not least, I would like to thank to my family – my parents, Weigang Zhou and Qianling Zhang, who spent the second half of their lives raising me and educating me to pursue my dreams, albeit at the expense of their own; and my lovely wife, Xiaodong Zhang, who has accompanied me, supported me, and endured me along our journey in the past 8 years. They are my fountain of courage and strength that brings me where I am and where I will go tomorrow.

Preface

The great master of warfare in the ancient China, Bin Sun, once said,

> Between heaven and earth, nothing is as noble as humanity… The right seasonal timing, the advantages of the terrain, the harmony among personnel – if these three things are not gained, there is calamity even in victory.
>
> (*Timing combat, the Art of Warfare, Sun Bin*)
>
> “间於天地之间，莫贵於人。…天时，地利，人和，三者不得，虽胜有殃。”— 孙膑，《月战•孙膑兵法》

Distilled from his military experience during the Warring state period of China, Sun Bin summarized that soldiers, as the main entity of the war, was one of the most important elements in succeeding a battle, and so did other factors related the natural environment. Leading army to campaign in the right seasoning and combating with advantages of the terrain could reduce the loss of productive labor, minimize the environmental destruction and prevent excessively high casualty rates. These benefits would build up sustainable victories in the period of warfare. His mind highlighted that not only the intrinsic elements, the individuals fighting the war, but also the extrinsic elements, the timing and terrain, determined the outcome of a battle. Biological processes seem to follow the same rule: they are driven by both the intrinsic and extrinsic factors. In my voyage of exploring the secretes of life, seeking how the intrinsic and extrinsic factors dictate the outcomes of biological processes has led to my investigation of the mechanisms of transcriptional control.

Chapter I

Introduction

to

Eukaryotic Transcriptional Control

1. Transcriptional control

In 1958, the central dogma was proposed to describe the basic principles of molecular biology in living organisms (Crick, 1970; Crick, 1958). A critical component of the central dogma is transcription, the process by which the genetic information encoded in DNA is transferred to its information carrier, mRNA. With the work of generations of scientists, we now understand many aspects of the transcription process inside cells. Transcription controls the expression of the genome in specific cell types, at different stages of the life cycle, and in response to a variety of environmental signals. Specific regulation of transcription is essential for the precise control of many biological processes, for instance, embryonic development, immune responses, neuronal plasticity, and survival under extreme environments. Alterations in the process of transcriptional regulation are commonly associated with human diseases and disorders (Jimenez-Sanchez et al., 2001; Vaquerizas et al., 2009).

The regulation of gene transcription in eukaryotic cells takes place at different steps in the process: initiation, elongation and the termination of transcription, and post-transcriptional processing. One of the major regulatory steps is the binding of transcription factors to DNA to activate gene transcription (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997). intrinsic factor, the DNA binding sequences, and extrinsic factors, the other DNA binding and regulatory factors can influence the binding and regulation of transcription factors.

2. The intrinsic factor: Specific DNA binding sequences

Transcription factors are a class of proteins that bind to DNA sequences and regulate the process of gene transcription (Latchman, 1997). A transcription factor usually contains one or more DNA binding domains. The binding of transcription factors to DNA typically requires recognition of specific DNA sequences (Jacob and Monod, 1961; Mitchell and Tjian, 1989; Ptashne and Gann, 1997). These sequences usually consist of 4-8 nucleotides (Badis et al., 2009; Badis et al., 2008; Harbison et al., 2004; Newburger and Bulyk, 2009; Wei et al., 2010; Zhu et al., 2009), with some rare exceptions that can span over 10 base pairs for an individual transcription factor (Whittle et al., 2009). The base-pair specific interactions between DNA sequences and the DNA binding domains provide the intrinsic basis for sequence-specific recognition, a fundamental aspect of gene-based transcriptional control.

Inside cells, sequence variation in the DNA binding sequences affect transcription factor binding occupancy and gene expression (Kasowski et al., 2010; Schmidt et al., 2010; Zheng et al., 2010). Kasowski and colleagues examined the binding profile of nuclear factor κB (NF-κB) among several human individuals and chimpanzees, where they observed that single nucleotide polymorphisms at the binding sequences of NF-κB were frequently associated with differences in NF-κB binding (Kasowski et al., 2010). Similarly, single-nucleotide differences in the Ste12 binding sites explained the majority of the differential Ste12 binding and regulation among 43 segregants of a cross between two distally related yeast species (Zheng et al., 2010). These studies highlighted the causal relationship between the binding sequences of a transcription factor and its binding and regulation *in vivo*.

However, the binding sequences of transcription factors may not be sufficient to direct binding *in vivo*. In the same study conducted by Kasowski and colleagues, while the polymorphisms at NF-κB binding sites were associated with changes in NF-κB binding occupancy, this genetic difference only explained 35% of the total variance in transcription factor binding among human individuals (Kasowski et al., 2010). In another study, the binding of CCAAT/enhancer-binding protein alpha (CEBPA) was examined in liver tissues of five vertebrates spanning from chicken to human (Schmidt et al., 2010). The binding profile of CEBPA displayed a strong species-specific pattern despite similarity in the consensus binding sequences of CEBPA (Schmidt et al., 2010), suggesting that factors in addition to the DNA sequences significantly influence transcription factor binding. In line with this idea, it is commonly observed that the high affinity binding sequences occur far more frequently than the number of binding events detected experimentally, even for transcription factors in organisms with a relatively small genome such as *Saccharomyces cerevisiae* (Harbison et al., 2004; MacIsaac et al., 2006). Therefore, the intrinsic binding affinity of a transcription factor may only be the tip of iceberg of the mechanisms controlling gene-specific transcription. Understanding how extrinsic factors contribute to the selection of genes bound and/or regulated by transcription factors is critical to revealing the molecular mechanisms of the specific transcriptional control.

3. The extrinsic factors:

chromatin, competition, and more

The binding sequences of transcription factors are embedded within the nuclear genome. Many components in the nucleus, such as chromatin and other DNA binding factors, may influence *in viv*o binding and regulation of transcription factors.

In eukaryotic cells, linear genome is organized into the chromatin, a structure comprised of DNA and histones. The basic packing unit of chromatin is the nucleosome, consisting of a segment of DNA wound around a core histone octamer. A string of nucleosomes can be assembled into a helical nano-fiber with diameter of ~30 nm, and this nano-fiber can be further compacted into a higher order structure forming heterochromatin, where it becomes inaccessible to the basic transcription machinery. Due to the condensed structure of heterochromatin and strong interaction between histones and DNA, chromatin has been suggested to limit the accessibility of binding sites to transcription factors (Khorasanizadeh, 2004; Kornberg and Lorch, 1999; Narlikar et al., 2002). This restriction could happen at two levels - at the individual nucleosome level, the association of DNA with the core histone complex may prevent the binding of transcription factors to the sites within the wound DNA; at the chromatin level, the heterochromatin may limit the regional accessibility to most sequence-specific transcription factors. These hypotheses are beginning to be examined with the advent of genomic tools such genome-tiling microarrays and high throughput sequencing that can map the positions of nucleosomes across the genome (Lee et al., 2007; Shivaswamy et al., 2008; Yuan et al., 2005). Studies have shown that the overall promoter nucleosome occupancy restricts the binding of transcription factor Leu3 (Liu et al., 2006; Wasson and Hartemink, 2009). However, it is still

unclear how and to what degree the individual nucleosome occupancy influences the sequence-specific binding of transcription factors. Heterochromatin is limited in the model system of budding yeast, and the role that heterochromatin might play in this process remains mysterious.

Other transcription factors, such as cooperating and competing factors, have also been shown to influence transcription factor binding (Pan et al., 2010; Pierce et al., 2003). At the *MYC* promoter, the binding site of ETS1 overlaps with the binding site of E2F, and it was only after mutation of the E2F binding site that ETS1 bound to the *MYC* promoter *in vivo* (Albert et al., 2001), suggesting that the binding of E2F prevents the binding of ETS1. Similarly, when budding yeast transits into meiosis, the transcriptional activator Ndt80 competes with a transcriptional repressor Sum1 to determine the expression of a set of sporulation genes, through competitive binding at their partially overlapped binding sequences (Pierce et al., 2003). These studies serve as examples where the competitive binding between transcription factors influences the transcriptional regulation of several loci. On the other hand, Pan and colleagues proposed three types of co-regulation between multiple transcription factors in collaboratively recognizing regulatory elements, based on the order of transcription initiation events (Pan et al., 2010). Overall, these studies suggested that *trans* influences might play a critical role in determining the specific binding of transcription factors. Nevertheless, there is still a lack of systematic analysis to address their influence on the specificity of transcriptional control,

An additional layer of complexity exists in the discrepancy between transcription factor binding and its subsequent gene regulation. Binding of a transcription factor *per se* is often not sufficient to activate or repress gene transcription (Birney et al., 2007; Farnham, 2009; Harbison et al., 2004; MacIsaac et al., 2006). For instance, in *Saccharomyces cerevisiae*, the number of detected binding events of transcription factors is 10 to 20 times more than the number of genes

regulated by these factors (Harbison et al., 2004). During early T-cell lineage commitment, several key transcriptional regulators are bound to over 30,000 binding sites but fewer than 4,000 genes exhibit differential expression along the process of T-cell commitment (Chen et al., 2012). The connection between the identified binding events of transcription factors and their capability to activate or repress gene expression remains a challenging issue to understand eukaryotic transcriptional programs.

In addition, recent studies showed the transcription of a gene can be influenced by several other mechanisms as well. The interactions within and between chromosomes present a dynamic and complex picture of eukaryotic genome in three-dimensions and such chromosomal interactions may play an important role in regulating gene expression (Dekker, 2008). Spatial localization of gene sequences and compartmentalization of molecular machinery within the nucleus may influence gene expression (Schneider and Grosschedl, 2007). Binding of transcription factor may release pulsed RNA polymerase at the promoter and may add another level of complexity immediately after transcriptional initiation (Rahl et al., 2010). The extent to which these mechanisms affect the specificity of transcription factors awaits further investigation.

In summary, it is still unclear what determines the genomic locations to which a transcription factor binds and whether this binding is able to influence the transcription of a gene (Farnham, 2009). To further complicate matters, transcription factors of the same family contain structurally conserved DNA binding domains and usually recognize similar short DNA motifs (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). However, these factors frequently regulate distinct biological responses (Robinson and Lopes, 2000; Sharrocks, 2001). The conflict between similarity in binding and divergence in regulation raises two questions: How are

distinct patterns of transcriptional regulation achieved? How is the regulation by a given transcription factor influenced by others that recognize very similar DNA motifs? To answer these questions, it is necessary to systematically interrogate, on a genome-wide scale, the factors that may contribute to the specificity of transcription factor binding and regulation.

4. References

Albert, T., Wells, J., Funk, J.O., Pullner, A., Raschke, E.E., Stelzer, G., Meisterernst, M., Farnham, P.J., and Eick, D. (2001). The chromatin structure of the dual c-myc promoter P1/P2 is regulated by separate elements. J Biol Chem *276*, 20482-20490.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Chen, Y., Negre, N., Li, Q., Mieczkowska, J.O., Slattery, M., Liu, T., Zhang, Y., Kim, T.K., He, H.H., Zieba, J.*, et al.* (2012). Systematic evaluation of factors influencing ChIP-seq fidelity. Nat Methods.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology *12*, 138-163.

Dekker, J. (2008). Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. J Biol Chem *283*, 34532-34540.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. Nat Rev Genet *10*, 605-616.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J.*, et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev *17*, 1309-1320.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Jimenez-Sanchez, G., Childs, B., and Valle, D. (2001). Human disease genes. Nature *409*, 853-855.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E.*, et al.* (2010). Variation in transcription factor binding among humans. Science *328*, 232-235.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259-272.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Latchman, D.S. (1997). Transcription factors: an overview. The international journal of biochemistry & cell biology *29*, 1305-1312.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D., and Lieb, J.D. (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res *16*, 1517-1528.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Newburger, D.E., and Bulyk, M.L. (2009). UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. Nucleic Acids Res *37*, D77-82.

Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends Genet *26*, 75-83.

Pierce, M., Benjamin, K.R., Montano, S.P., Georgiadis, M.M., Winter, E., and Vershon, A.K. (2003). Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. Mol Cell Biol *23*, 4814-4825.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Robinson, K.A., and Lopes, J.M. (2000). SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res *28*, 1499-1505.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S.*, et al.* (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036-1040.

Schneider, R., and Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. Genes Dev *21*, 3027-3043.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat Rev Mol Cell Biol *2*, 827-837.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol *6*, e65.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat Rev Genet *10*, 252-263.

Wasson, T., and Hartemink, A.J. (2009). An ensemble model of competitive multi-factor binding of the genome. Genome Res *19*, 2101-2112.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J *29*, 2147-2160.

Whittle, C.M., Lazakovitch, E., Gronostajski, R.M., and Lieb, J.D. (2009). DNA-binding specificity and in vivo targets of Caenorhabditis elegans nuclear factor I. Proc Natl Acad Sci U S A *106*, 12049-12054.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., and Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. Nature *464*, 1187-1191.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M.*, et al.* (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res *19*, 556-566.

Chapter II

Mechanisms of Transcriptional Regulation in the PHO Pathway

1. Introduction

1.1. The PHO pathway in *Saccharomyces cerevisiae*

Phosphorus is one the most abundant elements in living organisms. A chemical form of phosphorus, inorganic phosphate, is an essential building block for the synthesis of nucleic acids, phospholipids, and a variety of cellular metabolites. The requirement for phosphorus in the budding yeast *Saccharomyces cerevisiae* is met by the uptake of inorganic phosphate from the environment. Millimolar concentrations of inorganic phosphate are required for the synthesis of these molecules inside the cells (Auesukaree et al., 2004), whereas the concentrations of phosphate are considerably lower in the natural environments. Therefore, yeast cells have developed active strategies to regulate their uptake and utilization of inorganic phosphate according to its availability, and to counteract the changes in the concentrations of environmental phosphate. This process is controlled by the phosphate-responsive signaling (PHO) pathway in the budding yeast (Lenburg and O'Shea, 1996).

When budding yeast grows under phosphate depleted conditions, the PHO pathway specifically induces the expression of a set of genes (PHO genes) that function in inorganic phosphate transport (phosphate transporters), phosphate storage (synthesis and degradation of polyphosphate) and phosphate regeneration (phosphatases). When budding yeast is grown under phosphate replete conditions, the expression of these genes is strongly repressed (Oshima, 1997), suggesting a tight and specific control of PHO gene expression. The transcription of these PHO genes is controlled by two DNA binding proteins – Pho2 and Pho4 (Vogel et al., 1989). Pho2 is a pleiotropic transcription regulator; it also regulates genes in biosynthetic pathways of purine nucleotides and histidine (Arndt et al., 1987; Daignan-Fornier and Fink, 1992; Tice-Baldwin et

al., 1989), in addition to the PHO genes. Pho4 is the specific activator of the phosphate starvation response. The ability of Pho4 to activate gene transcription is regulated by its nuclear localization, which is controlled by the cyclin/cyclin-dependent kinase complex Pho85-Pho80 (Kaffman et al., 1994; Komeili and O'Shea, 1999; O'Neill et al., 1996; Schneider et al., 1994). Several other components contribute to the regulation of Pho4 localization as well: Msn5, a member of the β–importin family of nuclear transport receptors, exports Pho4 from the nucleus to the cytoplasm (Kaffman et al., 1998a); Pse1, another member of the β–importin family, imports Pho4 from the cytoplasm into the nucleus (Kaffman et al., 1998b); Pho81, an inhibitor of cyclin-dependent kinase, regulates the activity of the Pho85-Pho80 complex (Schneider et al., 1994).

The cellular localization of Pho4 is regulated in response to phosphate starvation (Figure 1). In high phosphate conditions, unphosphorylated Pho4 in the cytoplasm is recognized by Pse1 and is transported into the nucleus (Kaffman et al., 1998b). The nuclear Pho85-Pho80 complex phosphorylates Pho4 at multiple serine residues (Kaffman et al., 1994). Two of these phosphorylated sites are recognized by Msn5, which exports the phosphorylated Pho4 back into the cytoplasm (Komeili and O'Shea, 1999). Phosphorylated Pho4 remains in the cytoplasm until it is dephosphorylated, because the phosphorylation of Pho4 inhibits the interaction between Pho4 and its transporter Pse1 (Komeili and O'Shea, 1999). Therefore, the phosphorylation of Pho4 depletes Pho4 from the nucleus and the expression of the PHO genes is turned off. In low phosphate conditions, the decrease in phosphate concentration signals through a phosphate metabolite, inositol pyrophosphates (IP_7), which mediates the Pho81-dependent inhibition of Pho85-Pho80 cyclin-cyclin-dependent kinase complex (Lee et al., 2007b). The increase in IP_7 level in low phosphate conditions inhibits the kinase activity of Pho85-Pho80 complex and shuts

down the phosphorylation of Pho4 (Lee et al., 2007b). Unphosphorylated Pho4 continues to be shuttled into the nucleus by Pse1 but is no longer transported back to the cytoplasm by Msn5, owing to the lack of recognition between Msn5 and the unphosphorylated Pho4. As a result, Pho4 accumulates within the nucleus and is able to induce transcription of the PHO genes.

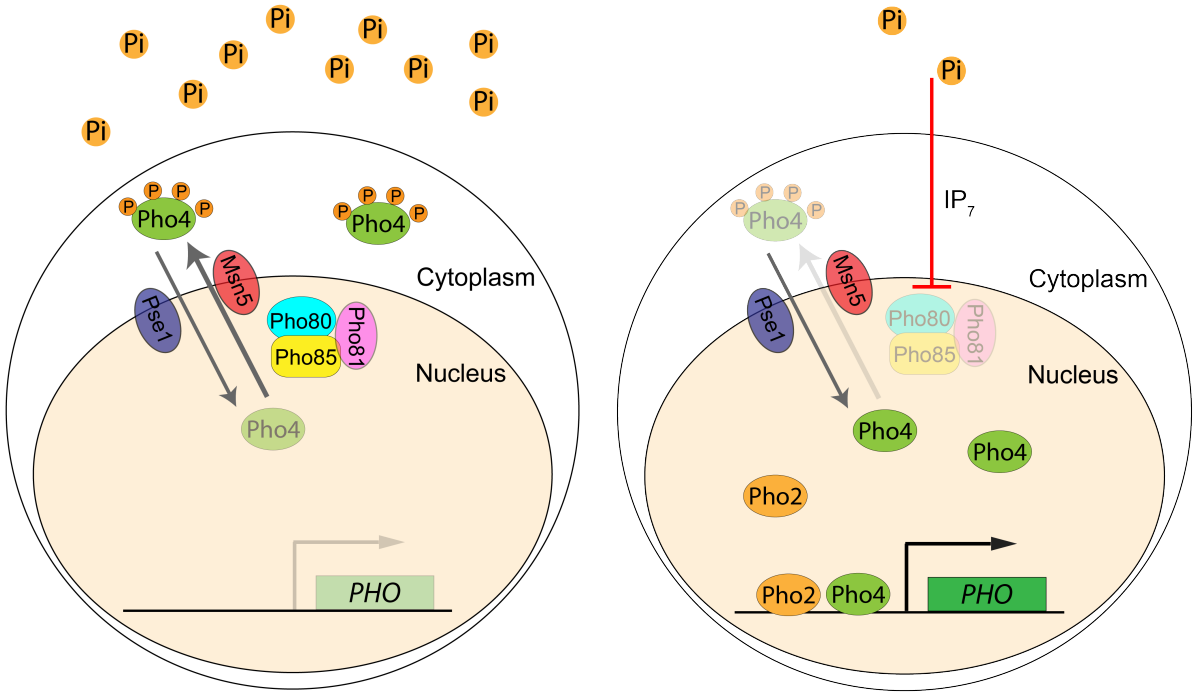


Figure 1. Phosphate-responsive signaling pathway in high phosphate conditions (left) and low phosphate conditions (right) in *S. cerevisiae*.

1.2. Pho4 and bHLH transcription factors

Sequence-specific transcription factors usually recognize DNA through base-pair specific interactions made by a DNA binding domain (Mitchell and Tjian, 1989; Ptashne and Gann, 1997). DNA binding domains can be categorized into different families based on their similarities in protein structure. Pho4, the key transcriptional regulator of the PHO pathway in *S. cerevisiae*, belongs to the basic helix-loop-helix (bHLH) family, which is the third largest transcription factor family in eukaryotic cells and regulates various cellular responses, ranging from cell proliferation, development, immune response, to basic cellular metabolism (Massari and Murre, 2000).

The regulatory elements of bHLH transcription factors share a signature DNA sequence motif, a core hexanucleotide sequence 'CANNTG' named the E-box sequence (Ephrussi et al., 1985). The structure of the basic HLH domain in complex with DNA revealed a number of features of how the bHLH transcription factors specifically bound to its recognition sequences (Ellenberger et al., 1994; Ferre-D'Amare et al., 1994; Ferre-D'Amare et al., 1993; Ma et al., 1994; Shimizu et al., 1997). The basic region of the bHLH domain makes nucleotide-specific contacts in the major groove of DNA (Massari and Murre, 2000; Shimizu et al., 1997), contributing to the recognition of E-box sequence (Figure 2). The second helix of the bHLH domain stabilizes the dimerization between two bHLH proteins through van der Waals interactions (Massari and Murre, 2000). Pho4 binds to DNA as a homo-dimer and recognizes a palindromic sequence, where the basic region of each bHLH domain recognizes half of its E-box sequence (Figure 2) (Shimizu et al., 1997).

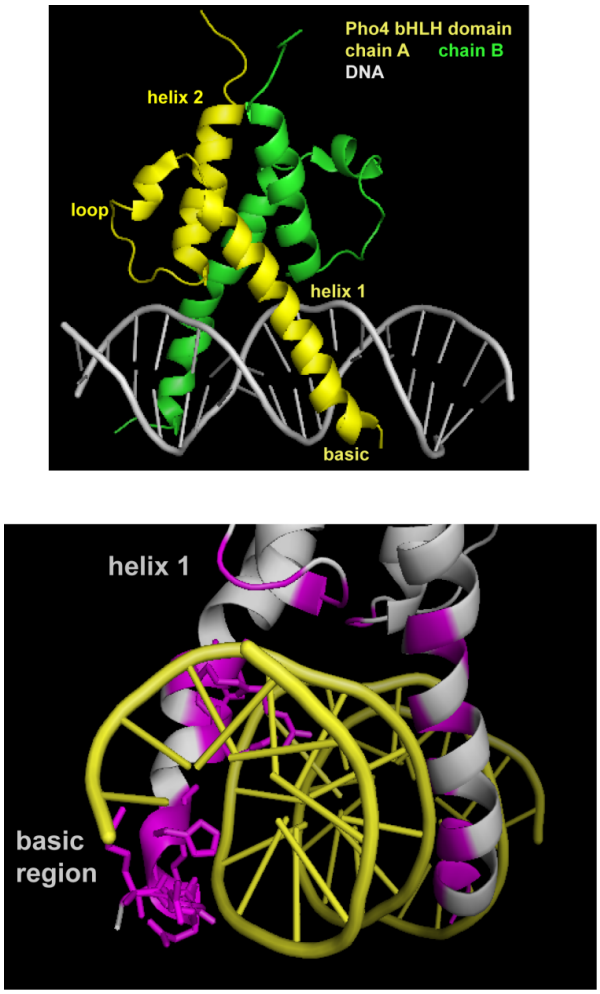


Figure 2. The structure of Pho4 bHLH DNA binding domain in complex with double strand DNA (PDB 1A0A) (Shimizu et al., 1997).

The upper panel shows the crystal structure of the homo-dimmer of Pho4 bHLH domains in complex with DNA. The two monomers of bHLH domain are colored in yellow and green respectively. The basic region of the first helix extends into the major groove of double strand DNA. The second helix mediates the interaction between the two monomers. The lower panel presents the sequence specific interaction between the basic region of bHLH domain and the major groove of DNA. The residues making contacts with DNA are colored in pink and shown wit their side chains.

1.3 Specific transcriptional control in the PHO pathway

The PHO pathway has proved to be a remarkable model system to study transcriptional activation in eukaryotic cells (Weake and Workman, 2010). Numerous studies provided molecular details describing how Pho4 activates the transcription of *PHO5*, a gene encoding a secreted phosphatase in response to phosphate starvation. Two upstream *cis* regulatory elements are critical in this process – 'CACGTG', one E-box sequence with high affinity for Pho4, and 'CACGTT', a sequence with lower affinity (Lam et al., 2008). It has been demonstrated that the arrangement of these Pho4 binding sites and nucleosome plays an important role in determining the dynamics and the level of transcriptional activation (Lam et al., 2008; Venter et al., 1994), suggesting that nucleosomes influence the binding and function of Pho4. In addition, both Pho2 and Pho4 are required for the activation of *PHO5*, and both factors bind cooperatively to the promoter of *PHO5* through a direct interaction (Barbaric et al., 1996; Vogel et al., 1989). After binding to DNA, these factors recruit chromatin remodelers and subunits of the transcription initiation complex to activate gene transcription (Svaren and Horz, 1997).

Although much is known regarding how Pho4 activates the transcription of *PHO5*, little is understood about why Pho4 activates *PHO5* and a few other PHO genes. Answering this question requires systemic examination of all the Pho4-regulated genes and identification of their common regulatory elements. Two previous studies characterized the number of genes being induced by Pho4 in response to phosphate starvation (Ogawa et al., 2000; Springer et al., 2003). 19 and 20 genes were identified to be Pho4-dependent respectively, with 14 genes overlapping between the two studies. All of these 14 genes carry at least one 'CACGTG' hexanucleotide

sequence in their promoters, suggesting that this sequence motif is indeed the primary *cis*-regulatory element of Pho4. *In vitro* studies also confirmed that the DNA binding domain of Pho4 specifically recognizes 'CACGTG' sequences with high affinity (Badis et al., 2008; Maerkl and Quake, 2007; Zhu et al., 2009a). It is thus likely that this 6-mer sequence confers the specificity required for Pho4 to selectively regulate the PHO genes. However, the information encoded within the hexanucleotide sequence tells otherwise. Each position of the hexanucleotide sequence provides 4-bit information to specify the sequence from a pool of random DNA sequences. The 6-mer DNA sequence encodes up to 4^6 bit information, which is sufficient to select one specific sequence out of 4000 bp DNA. By this calculation, roughly 3000 'CACGTG' sequences exist randomly in the 12 million base pair genome of *S. cerevisiae*, hundreds of times higher than the number of genes being regulated by Pho4.

How does Pho4 selectively regulates only a handful of PHO genes out of ~6000 genes in the budding yeast genome, where many other genes could carry the same 'CACGTG' binding sequences at their promoters? Here I intended to answer this question by combining several genomic approaches to systematically dissect how Pho4 specifically recognizes its binding sites and regulates gene transcription. First, I examined the determinants of Pho4 binding at the whole genome scale. Is the intrinsic binding affinity of Pho4 necessary and sufficient to determine Pho4 binding *in vivo*? If not, what are the other factors that contribute to the specific binding pattern of Pho4? Secondly, I investigated the connection between Pho4 binding and its ability to activate the transcription of a gene. How many of the Pho4 binding events are functional? If not all, what are the underlying mechanisms that goven the functionality of transcription factor binding? Last, I constructed computational models to predict the binding and

regulation pattern of Pho4, in order to assess our understanding of the determinants for the specific control of PHO pathway.

2. Results

2.1. *in vitro* binding of the transcription factor Pho4

The *in vitro* binding specificity of Pho4 has been carefully examined by high throughput biophysical methods, such as protein-DNA binding microarrays (PBM) (Badis et al., 2008; Zhu et al., 2009a) and mechanically induced trapping of molecular interactions (MITOMI) (Maerkl and Quake, 2007). A consensus hexanucleotide E-box sequence, 'CACGTG', was identified from these studies as being bound by Pho4 with high affinity, consistent with the atomic structure of Pho4 in complex with double stranded DNA (Shimizu et al., 1997). This hexanucleotide sequence was also found in the promoters of most of Pho4-regulated genes, including *PHO5*, *PHO84*, *SPL2*, *PHO8* etc (Ogawa et al., 2000), and was confirmed as the critical element of Pho4-dependent transcriptional activation (Oshima, 1997). Therefore, it appeared that the *in vitro* binding preference of Pho4 represented, at least partially, the *in vivo* binding preference.

The above affinity measurements provided a quantitative description of the intrinsic binding preference of Pho4 to different DNA sequences, which can be transformed into a Position Specific Scoring Matrix (PSSM) or Position Weighing Matrix (PWM) (Stormo and Zhao, 2010). The PSSM of Pho4 (Table 1) renders a non-biased way to identify the DNA sequences with highest affinity for Pho4 across the budding yeast genome. To understand the extent to which specific DNA sequence recognition mediates *in vivo* binding of Pho4, and whether additional DNA sequence elements contribute to the binding site selection of Pho4, I first examined whether the *in vitro* identified high affinity binding site for Pho4 is an important determinant of Pho4 binding *in vivo*.

Table 1. Position Specific Scoring Matrix of Pho4 binding *in vitro*.

The PSSM is derived from the data kindly provided by Sebastian Maerkl from his work (Maerkl and Quake, 2007). The PSSM is derived according to the method described in Lam et al (Lam et al., 2008). The numbers in the matrix represent the frequency of observing the indicate nucleotide (raw) at the specified position (column) of 8-mer DNA sequences that are bound by Pho4. The most favored nucleotide at each position is highlighted in bold.

	Position in the DNA sequence							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	0.244787	0.119955	**0.772446**	0.041627	0.054081	0.058688	0.055140	0.125808
C	0.255008	**0.759357**	0.050908	**0.844115**	0.060178	0.117959	0.065548	**0.374398**
G	**0.374398**	0.065548	0.117959	0.060178	**0.844115**	0.050908	**0.759357**	0.255008
T	0.125808	0.055140	0.058688	0.054081	0.041627	**0.772446**	0.119955	0.244787

2.2. The intrinsic affinity of Pho4 is an important determinant of Pho4 binding *in vivo*

Recently, the global binding pattern of Pho4 was determined with genome-wide Chromatin immunoprecipitation (ChIP) followed by microarray hybridization (ChIP-chip) and it revealed a broad range of in vivo binding events (Harbison et al., 2004; Nishizawa et al., 2008). Although the *in vitro* determined high affinity binding sequences of Pho4 are enriched in the set of sequences bound *in vivo* by Pho4,, many of the identified genomic regions did not contain a consensus Pho4 binding site (Harbison et al., 2004; Nishizawa et al., 2008), questioning the consistence between *in vivo* and *in vitro* binding of transcription factors. However, these results were complicated by low signal-to-noise ratio and potential biases from the experimental procedures (Ho et al., 2011; Peng et al., 2007; Waldminghaus and Skarstad, 2010). Moreover, the identified Pho4 binding events were rarely refined to a narrow genome region and it was therefore difficult to pinpoint the binding of Pho4 to a specific binding sequence. It remains unclear if Pho4 *in vitro* binding specificity is an important determinant of Pho4 binding *in vivo*.

To overcome these caveats of the traditional ChIP-chip approach, I applied biotin-tagging chromatin immunoprecipitation to capture Pho4-bound genomic DNA (Kolodziej et al., 2009; van Werven and Timmers, 2006) and combined it with Illumina sequencing technology (Bio-ChIP-Seq) to determine the *in vivo* binding landscape of Pho4 at base pair resolution (Figure 3). This method harnesses the power of high throughput deep sequencing and exploits one of the strongest non-covalent interactions in nature – the interaction between biotin and straptavidin (4×10^{-14} M). This method allows stringent wash conditions during the pull-down procedure to reduce the non-specific background binding and improve the signal-to-noise ratio (Kolodziej et al., 2009).

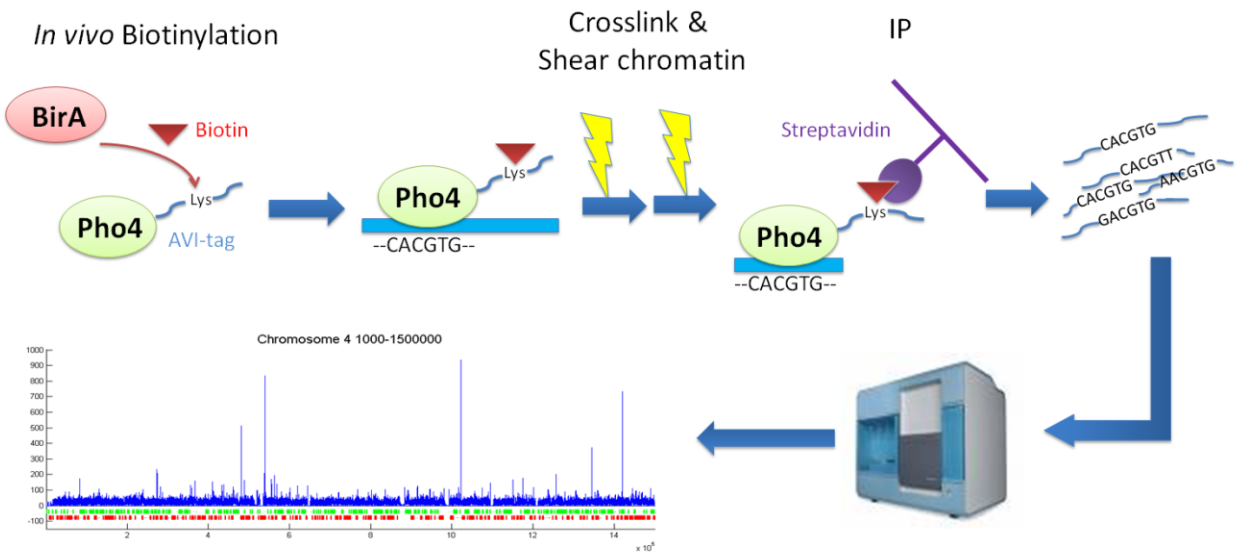


Figure 3. Schematic depicting the procedure of biotin tagging immunoprecipitation folloId by high throughput sequencing (Bio-ChIP-Seq).

AVI-tagged Pho4 under the control of its endogenous promoter was integrated into a strain containing the *E. coli birA* gene integrated at the *URA3* locus. BirA is fused with a nuclear localization sequence at its N-terminus. BirA recognizes the AVI-tag sequence and biotinylates the lysine residue within it. After chemical fixation with formaldehyde (1%), chromatin was sheared into small fragments and Pho4-bound DNA was selectively enriched through biotin-streptavidin affinity isolation and subjected to Illumina sequencing.

I examined Pho4 binding in both phosphate-replete (10 mM Pi) and phosphate starvation conditions (0 mM Pi), with genomic DNA input and mock-IP as controls. The genomic DNA input normalized non-uniform background due to DNA fragmentation for each ChIP experiment, and the mock-IP controlled for non-specific enrichment due to the experimental procedures. The DNA fragments of size between 50 bp and 150 bp were selected for sequencing, which gave rise to an estimated width of 300 bp for an individual Pho4 binding event. An example of Pho4 binding *in vivo* is shown in Figure 4 as a global view of the Pho4 occupancy on *S. cerevisiae* chromosome IV. Pho4 is specifically enriched at several genomic regions after phosphate starvation, visualized by the sharp peaks along the chromosome.

To systematically identify regions enriched in Pho4 binding, I adapted the detection algorithm of Peakseq (Rozowsky et al., 2009), which was initially used in identifying transcription factor binding in *C. elegans*. Briefly, I first identified putative binding regions with a false discovery rate of 0.05 estimated with a simulated random genome background. For each of these regions, I then calculated the statistical significance (P-value and Q-value) of Pho4 enrichment relative to its input and mock-IP after normalizing the depth of these sequencing data. With a conventional statistical significance of p-value 0.05, Pho4 binds to 108 distinct regions in the genome in phosphate starvation conditions (Appendix 3). These regions include the promoters of all the known *PHO* genes that are activated by Pho4 (Ogawa et al., 2000; Springer et al., 2003).

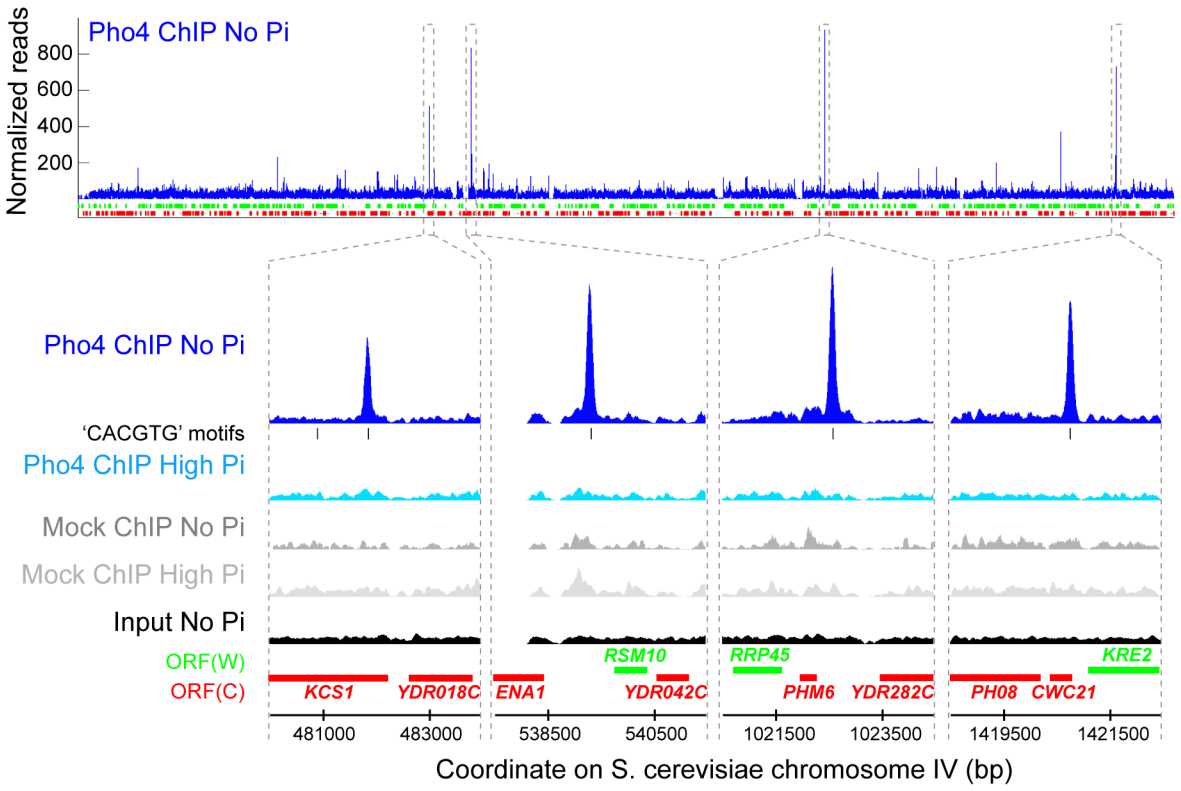


Figure 4. ChIP-seq analysis of Pho4 binding and mock IP on chromosome IV.

Pho4 binding was examined in *pho4::PHO4-AVI-TRP1* strain integrated with *E.coli birA* gene at *URA3* locus. Mock IP was performed with wild-type strain integrated with *birA*. The traces include Pho4 binding and Mock IP in high and no inorganic phosphate conditions, and genomic input in no phosphate conditions. All ChIP results were normalized to Pho4 ChIP at no Pi conditions.

I determined the high affinity Pho4 binding sequences using the PSSM derived from the *in vitro* binding affinity measurements of Pho4 (Lam et al., 2008; Maerkl and Quake, 2007). I selected the most stringent threshold (0.0075) to recapitulate *in vivo* validated high affinity binding sites as the threshold for the high affinity binding sequences (Lam et al., 2008). The resulting high affinity sequences contained 'CACGTG' as the core motif. All 'NCACGTGN' sequences except 'TCACGTGA' in *S. cerevisiae* genome meet the threshold. A second threshold of 0.0003 was chosen as the threshold of Pho4 low affinity binding sequences to represent the *in vivo* verified functional binding sites at the promoter of PHO genes (Barbaric et al., 1992; Munsterkotter et al., 2000; Ogawa et al., 1995; Venter et al., 1994). To circumvent the threshold effect in the binding event identification, I analyzed the presence of either high or low affinity binding sequences in the identified Pho4 binding regions at different statistical thresholds (Table 2). The *in vitro* DNA binding specificity of Pho4 explained at least 88% of the Pho4 binding regions, in contrast to 6 out of 108 as expected by chance.

Table 2. The number of identified genomic regions containing the high affinity and low affinity Pho4 binding sequences in *S. cerevisiae.*

The numbers in the parenthesis indicate the number of regions expected by chance.

Total	High affinity motifs	Low affinity motifs	Without motif	Threshold for binding	
				P-Value	Q-Value
108	66 (2)	29 (4)	13 (102)	0.05	-
95	65 (2)	23 (4)	7 (89)	0.01	-
64	55 (1)	9 (2)	0 (61)	-	0.05
55	48 (1)	7 (2)	0 (52)	-	0.01

To further test the possibility that the *in vitro* binding sequence of Pho4 is a critical determinant of Pho4 binding *in vivo*, I examined the position of these binding sequences relative to the position of Pho4 ChIP peaks in the identified binding regions. If Pho4 binding relies on its intrinsic affinity to DNA, the binding sequence should be beneath or close to the position of the ChIP binding peak. This is indeed what I observed for all the Pho4 binding regions that contain the high affinity binding sequence (Figure 5) or the low affinity sequences (data not shown). Together, I concluded that the intrinsic DNA binding affinity of Pho4 is an important determinant of Pho4 binding *in vivo*. In addition, this result also suggests that the interaction between Pho4 and other transcription factors (e.g. Pho2) does not significantly alter the DNA binding specificity of Pho4.

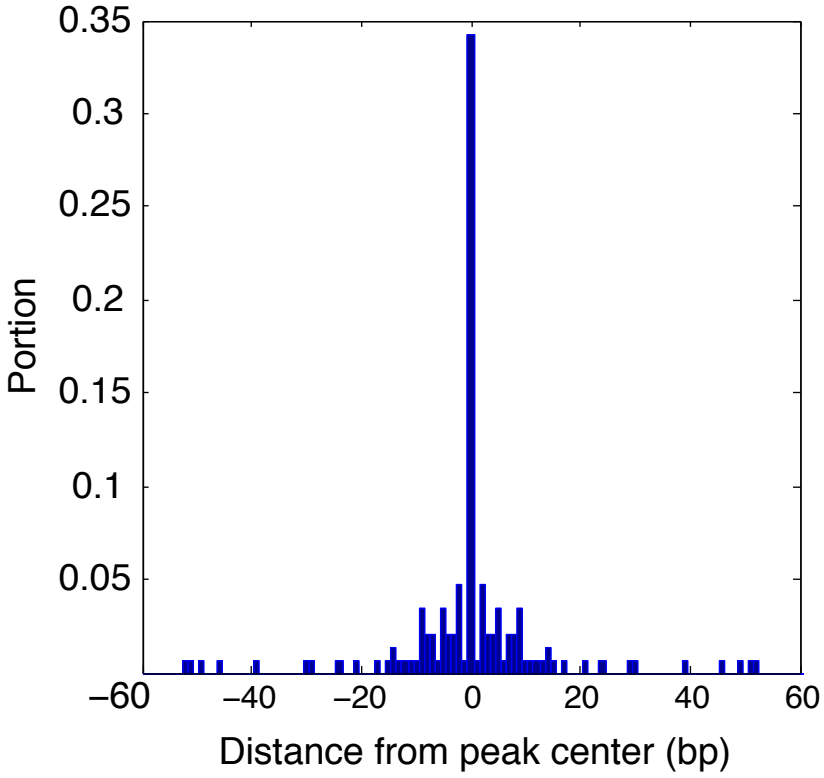


Figure 5. The peak-sequence distance for all the identified Pho4-bound regions.

The histogram shows the distribution of the distance between the position of the Pho4 binding sequence and the position of the Pho4 binding peak, for all the Pho4-bound regions that contain at least one high affinity binding site. The histogram remains the same for the regions that contain at least one low affinity binding site.

2.3. Binding sequences of Pho4 are not sufficient to determine Pho4 binding *in vivo*

Next, I sought to understand the extent of which the high affinity Pho4 binding motifs are able to specify Pho4 binding *in vivo*. Based on the PSSM score of 0.0075, there are 843 high affinity binding sites in *S. cerevisiae* genome. However, only a small portion of them are found within the Pho4 enriched ChIP regions. To exclude the possibility that the detection algorithm is not sensitive to Pho4 enriched locally at its binding sites, I developed a method to rigorously determine Pho4 binding at the level of its high affinity binding sites. In summary, I computed the local occupancy of Pho4 (20 bp window center at its binding sequence), genomic input and Mock-IP at every high affinity binding site and scored a signal as binding if it meets all of the following requirements: (1) significantly bound (ChIP occupancy, $p \leq 0.05$), (2) significantly enriched over input ($mean_{ChIP}/mean_{Input}$, $p \leq 0.05$), (3) significantly enriched over mock IP ($mean_{ChIP}/mean_{mock\ IP}$, $p \leq 0.05$). These statistical thresholds gave an overall p-value of 0.0066 estimated by randomly sampling the genome background.

Although there are 843 high affinity binding in the genome, I observed Pho4 binding to only 115 (~14%) of these sites (FDR = 0.05) (Figure 6), implying that factors other than DNA binding specificity influence Pho4 binding *in vivo*. Overall, the data suggest that the high affinity binding motifs of Pho4 are necessary but not sufficient to specify the *in vivo* binding landscape of Pho4.

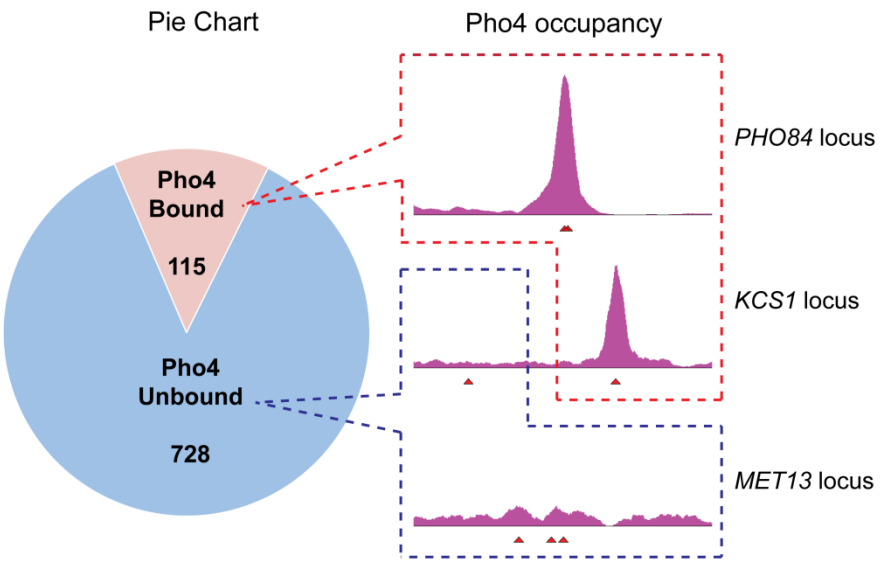


Figure 6. Pho4 binding at its high affinity binding sites.

On the left panel, the pie chart shows the portion of Pho4-bound and -unbound high affinity binding sites. On the right panel, examples of Pho4 occupancy at several loci are shown to illustrate the bound and unbound states of a high affinity binding site.

2.4. Chromatin negatively regulates Pho4 binding

The eukaryotic genome is packed into units of the nucleosome, a segment of DNA wound around a histone octomer complex. It has been suggested that nucleosomes might restrict the access of transcription factors to their potential binding sites (Khorasanizadeh, 2004; Kornberg and Lorch, 1999; Narlikar et al., 2002). To determine whether local chromatin structure influences the binding of Pho4 and if this restriction explains Pho4 binding pattern at the high affinity binding sites, I mapped nucleosome occupancy by isolating mono-nucleosomal DNA after micrococcal nuclease (MNase) digestion and paired-end Illumina sequencing (Figure 7).

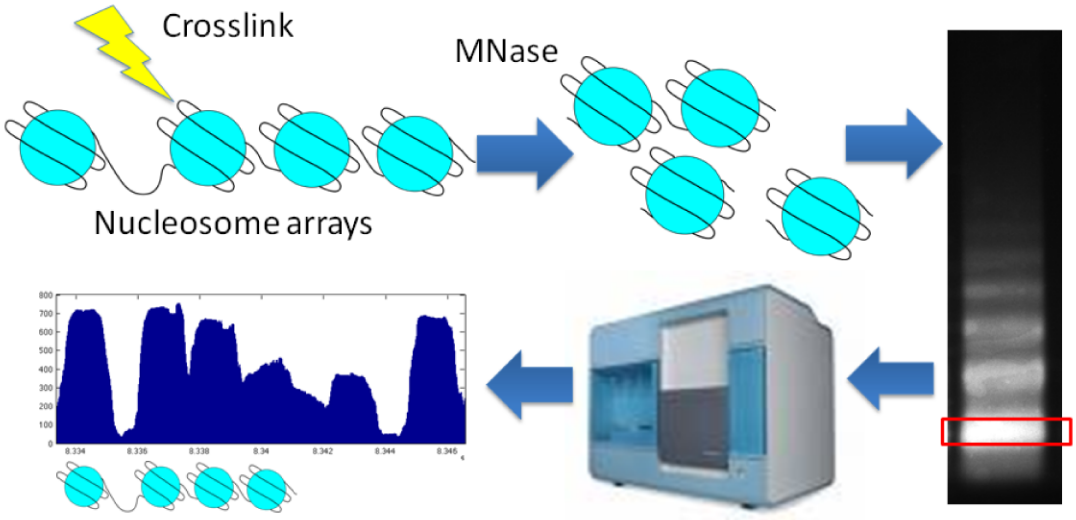


Figure 7. Schematic depicting the procedure of mapping *in vivo* nucleosome occupancy.

After chemical fixation with formaldehyde (1%), chromatin was isolated and subjected to Micrococcal nuclease digestion until most chromatin is digested to mono-nucleosome. The mono-nucleosomal NDA was visualized and purified from electrophoresis, and subjected to Illumina sequencing.

Paired-end DNA sequencing directly returns the precise location of the isolated mono-nucleosomal DNA, bypassing the sophisticated fitting algorithms needed to infer nucleosome positions for data with single-end sequencing (Weiner et al., 2010). Under the assumption that the center of each sequenced DNA fragment is the nucleosome dyad (Mavrich et al., 2008), I mapped the center of each sequenced nucleosomal DNA on the genome and extended 73 bp on each side to generate mono-nucleosome coverage (Figure 8). This assumption was confirmed by two lines of evidences: first, the length distribution of the sequenced nucleosomal DNA showed a periodicity of roughly 20-bp (Figure 9), consistent with the symmetrical progression of the MNase digestion every 10 bp on both ends (Kornberg and Lorch, 1999); second, the di-nucleotide frequency of the sequenced nucleosomal DNA showed a distinct feature of 10-bp periodicity after alignment at the fragment center (Figure 10), a hallmark of stable nucleosomes described in previous studies (Albert et al., 2007; Kaplan et al., 2009; Mavrich et al., 2008; Segal et al., 2006). Nucleosome occupancy maps are shown for the *PHO84*, *SPL2* and *GAL4* loci, under both phosphate-replete and phosphate starvation conditions (Figure 11).

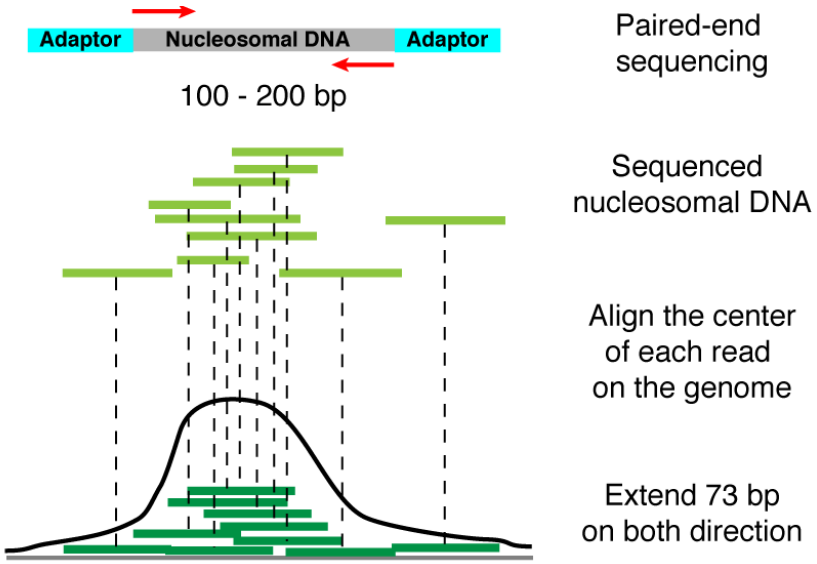


Figure 8. Schematic depicting alignment of sequencing reads to the *S. cerevisiae* genome.

The center of each paired-end sequenced nucleosomal DNA fragment with length between 100 to 200 bp is aligned on *S. cerevisiae* genome. Aligned reads are extended 73 bp in both directions from the center (nucleosome dyad).

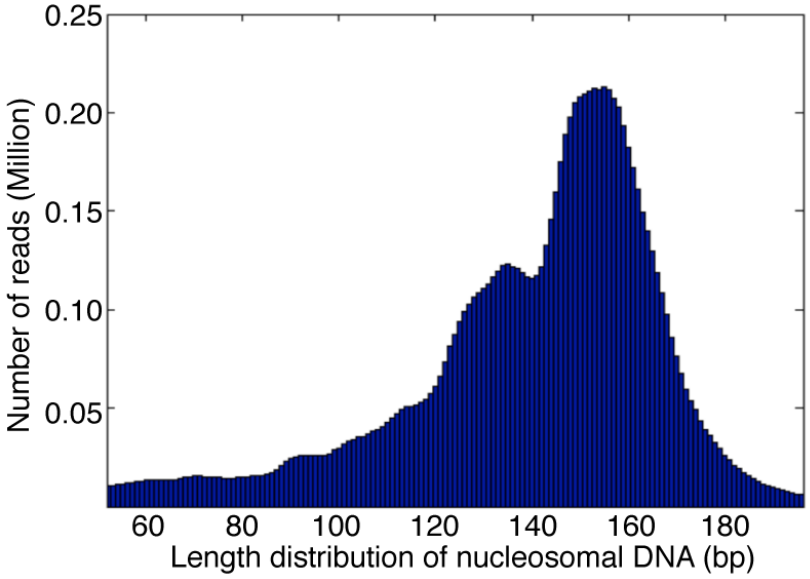


Figure 9. Length distribution of paired end sequenced nucleosomal DNA.

The major peak at ~150 bp reflects intact mono-nucleosomes (Albert et al., 2007; Mavrich et al., 2008). The distance between the major and minor peaks is ~20 bp, suggesting MNase digestion of 10 bp from both sides of the nucleosomes (Albert et al., 2007; Kaplan et al., 2009; Mavrich et al., 2008; Segal et al., 2006).

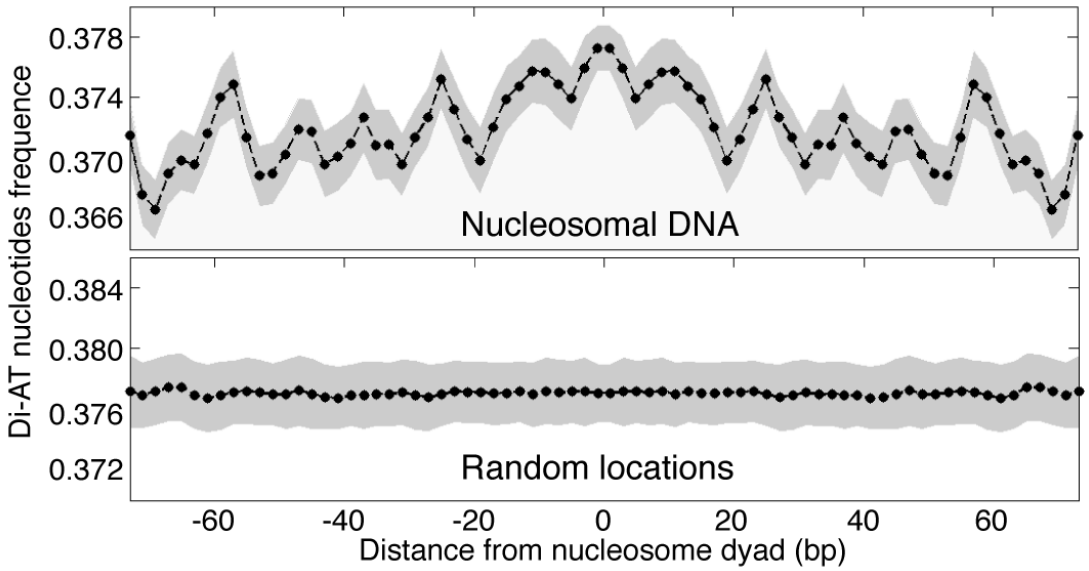


Figure 10. Di-AT (AA/AT/TA/TT) nucleotide frequency of sequenced nucleosomal DNA, averaged from 50 trials.

For each trial, the di-AT frequency is calculated from 50,000 randomly selected sequenced nucleosomal DNA fragments, assuming the center of each sequenced fragment is the nucleosome dyad (position 0). As a control, the same analysis was done for randomly selected genome locations. Black squares indicate the average and gray area denotes mean ± s.d. The ~10 bp periodic di-AT nucleotide frequency is similar to previously observed properties of nucleosomal DNA.

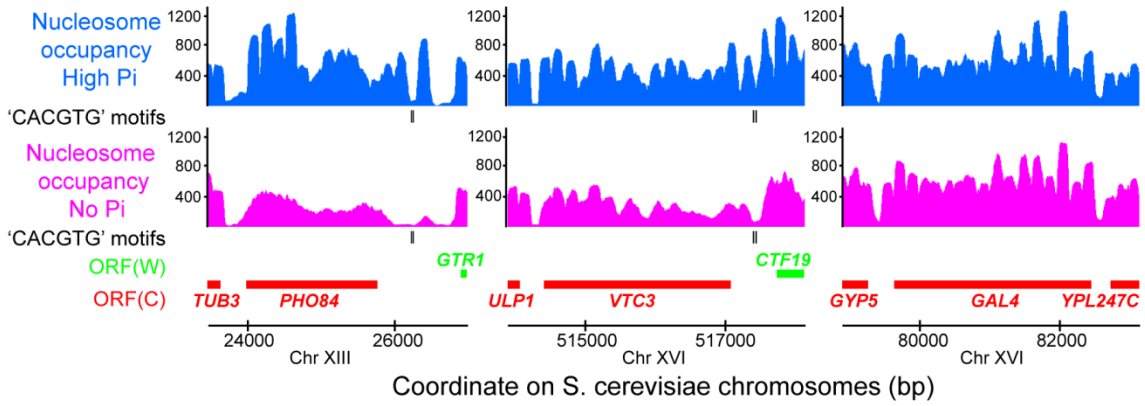


Figure 11. Examples of nulceosome occupancy map at both phosphate-replete and phosphate starvation conditions.

Nucleosome occupancy map for *PHO84* and *VTC3* (both de-repressed in no Pi conditions), *GAL4* (not responsive to Pi concentration) are shown in high and no Pi conditions. The nucleosome maps in both conditions Ire normalized to have the same number of total reads.

To examine the influence of nucleosome occupancy on the binding Pho4 *in vivo*, I calculated the average nucleosome occupancy of a 20-bp window centered on the high affinity binding sites and looked at the correlation between nucleosome occupancy and the occurrence of Pho4 binding. Chromatin negatively regulates Pho4 binding as Pho4 enrichment is generally observed at binding sites that have relatively low nucleosome occupancy (Figure 12).

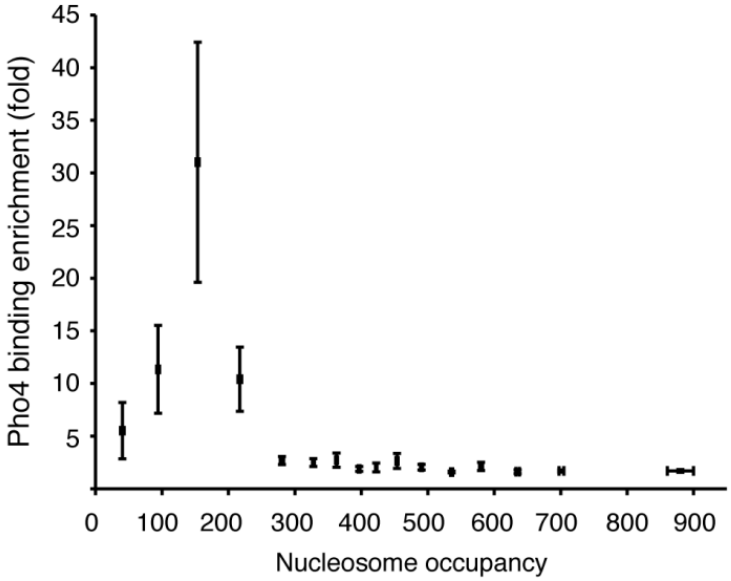


Figure 12. Pho4 binding enrichment and nucleosome occupancy in no Pi conditions.

All consensus sites were put into bin sizes of 50 and I ploted the average nucleosome occupancy against the average Pho4 binding enrichment for each bin.

To further quantify the restriction of nucleosome on Pho4 binding, I considered two possible states of a Pho4 high affinity binding site: an inaccessible state, occluded by nucleosomes; or an accessible state, exposed in a nucleosome-depleted region (nucleosome-free or nucleosome linker region). ~80% of the *S. cerevisiae* genome is estimated to be covered with nucleosomes (Lee et al., 2007a); I thus say a transcription factor binding site is in the accessible state if the average nucleosome occupancy on the site is in the lower quartile of the genome nucleosome occupancy. Symmetrically, I say the site is inaccessible if the average nucleosome occupancy is in the upper quartile of the genome nucleosome occupancy. As expected, Pho4 is not bound to the binding sites that are inaccessible and most occluded by nucleosomes (209 of 216 sites, 97%; Figure 13). However, using the same threshold, Pho4 is also not bound to two thirds of the most accessible sites (172 of 248 sites) (Figure 13). These two thresholds compartmentalized the analysis to the most accessible and least accessible consensus sites. To make sure the results above is not limited by the thresholds chosen for the accessible and inaccessible sites, I varied the percentile threshold over a wide range (15 % – 30%) and observed almost the same pattern of Pho4 binding at accessible and inaccessible sites (Figure 14). Therefore, I concluded that chromatin structure inhibits transcription factor binding, but open chromatin structure is not sufficient for Pho4 binding at its high affinity binding sites.

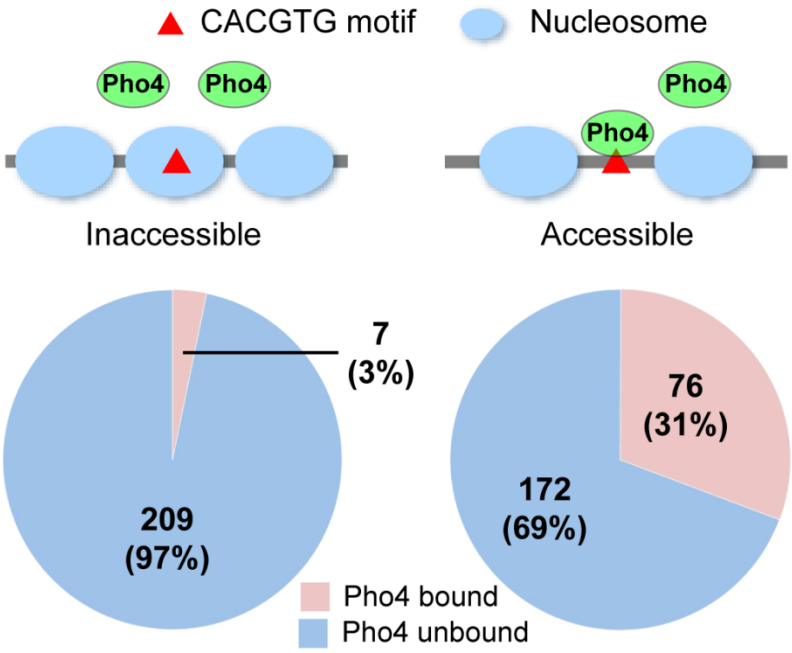


Figure 13. Pho4 binding at accessible and inaccessible high affinity binding sites.

Pie charts show Pho4 binding at inaccessible (the quartile most occluded by nucleosomes) and accessible (the quartile least occupied by nucleosomes) high affinity binding sites in no Pi conditions.

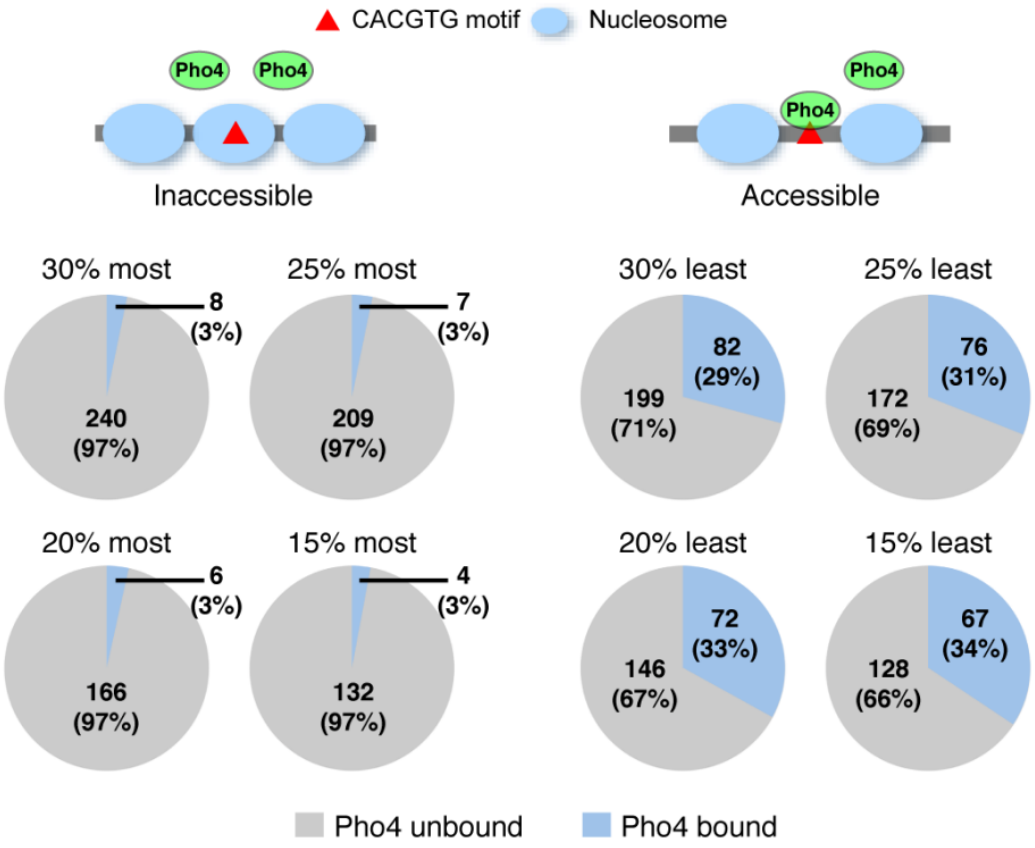


Figure 14. Pho4 binding at inaccessible and accessible high affinity binding sites in no Pi conditions.

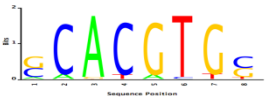
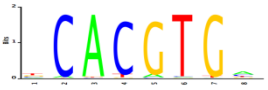
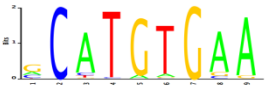
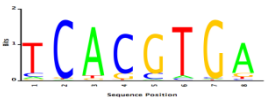
Pie charts show Pho4 binding at inaccessible binding sites that are most occluded by nucleosomes (left half) and at accessible binding sites that are least occupied by nucleosomes (right half). I observe similar results, independent of the fraction of inaccessible or accessible sites selected for analysis.

2.5. Competition from Cbf1 determines Pho4 occupancy *in vivo* at nucleosome depleted sites

The Pho4 high affinity binding sites share the same core motif 'CACGTG' and meet the minimum affinity threshold of which the site can be bound and regulated by Pho4 in response to phosphate starvation *in vivo*. Of the accessible sites that are not occluded by nucleosomes, why are some of these sites bound by Pho4 and the others are not? It is possible that proteins with similar specificity compete with Pho4 for binding to these accessible sites.

The DNA binding specificity of transcription factors is usually determined by the structure of their DNA binding domain; transcription factors with DNA binding domain belonging to the same structural family commonly recognize similar sequence motifs. Pho4 belongs to the basic Helix-Loop-Lelix (bHLH) transcription factor family, where members of this family recognize the E-box binding motif 'CANNTG' (Massari and Murre, 2000; Robinson and Lopes, 2000). I examined the DNA binding specificity of all known bHLH transcription factors in *S. cerevisiae*, and found out that three other factors also recognize 'CACGTG' sequences (Table 3). The three factors are: Cbf1, centromere binding factor 1, a dual function transcription factor regulating the sulfur starvation pathway and chromosome segregation (Cai and Davis, 1990; Mellor et al., 1991); Rtg3, ReTroGrade regulation 1, activating the retrograde and TOR pathways (Crespo et al., 2002; Jia et al., 1997; Rothermel et al., 1997); Tye7, transposable element Ty-mediated expression gene 7, activating the expression of glycolytic genes and Ty-mediated genes (Lohning and Ciriacy, 1994; Nishi et al., 1995; Sato et al., 1999).

Table 3. bHLH transcription factors that recognize 'CACGTG' motif *in vitro*

Transcription factors	DNA binding motif	Copies / cell (Ghaemmaghami et al., 2003)	Localization (Huh et al., 2003)
Pho4	 (Zhu et al., 2009a)	~1000*	Cytoplasm Nucleus
Cbf1	 (Zhu et al., 2009a)	6890	Nucleus
Rtg1	-	2190	Cytoplasm Nucleus
Rtg3	 (Zhu et al., 2009a)	1050	Cytoplasm Nucleus
Ino2	 (MacIsaac et al., 2006)	784	Nucleus
Ino4	 (MacIsaac et al., 2006)	521	Nucleus
Hms1	 (Morozov and Siggia, 2007)	-	-
Tye7	 (Zhu et al., 2009a)	486	Cytoplasm Nucleus

Among these bHLH transcription factors, Cbf1, is present in the nucleus at high concentration (Ghaemmaghami et al., 2003), is not known to interact with Pho4 (Graumann et al., 2004), and binds with high affinity to the same consensus binding motif 'CACGTG' *in vitro* and *in vivo* as does Pho4 (Harbison et al., 2004; MacIsaac et al., 2006; Maerkl and Quake, 2007; Zhu et al., 2009a). I therefore focused on examining the possibility that Cbf1 competes with Pho4 binding at some of these accessible 'CACGTG' sites. To test this hypothesis, I identified *in vivo* binding sites for Cbf1 in high and no Pi conditions.

In no Pi conditions, 77% (132 of 172) of the accessible high affinity binding sites not bound by Pho4 are occupied by Cbf1 (Figure 15). Intriguingly, most of the accessible sites bound by Pho4 (72 of 76, 95%) are also bound by Cbf1. Thus, the accessible, high affinity 'CACGTG' sites mainly fall into two classes: those where Cbf1 competes with Pho4 most effectively, resulting in detectable binding of Cbf1 but not Pho4; and a second class where Cbf1 competes less effectively, resulting in significant occupancy of both Pho4 and Cbf1. In high Pi conditions, Cbf1 is bound to both of these classes of sites (119 of 132 Cbf1-bound sites and 67 of 72 sites both occupied by Pho4 and Cbf1) (Figure 16). This is consistent with the fact that Cbf1's nuclear localization is independent of phosphate starvation. The observed patterns of Pho4 binding and Cbf1 binding at the accessible high affinity binding sites suggest that the binding of Cbf1 may prevent the binding of Pho4 and thus explains why Pho4 only occupies ~30% of its accessible sites.

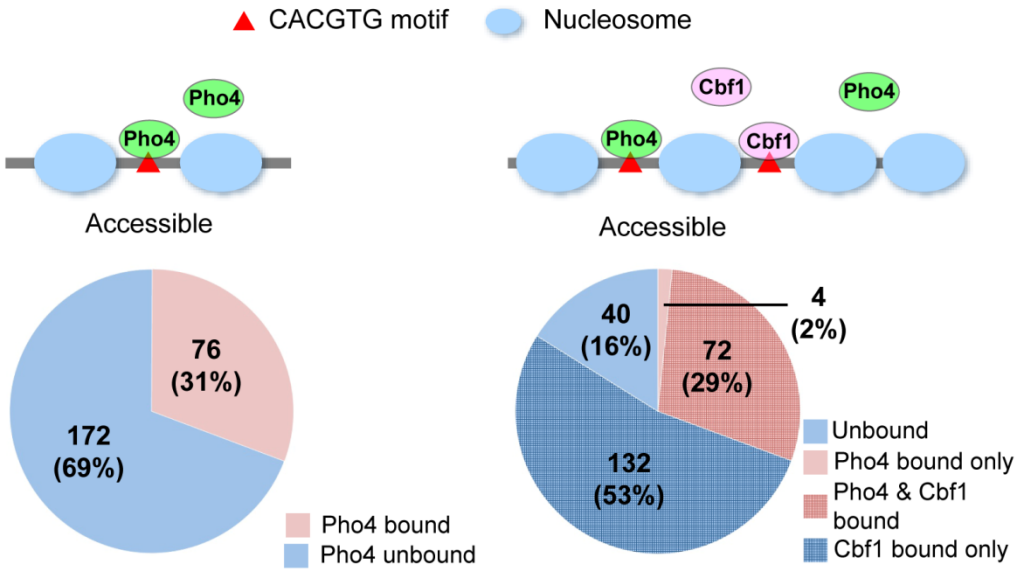


Figure 15. Pho4 and Cbf1 binding at the accessible sites in no Pi conditions.

On the left, pie chart shows Pho4 binding at the accessible high affinity binding sites in no Pi conditions. On the right, the pie chart is overlaid with the binding of Cbf1 with shadings. Cbf1 is considered bound to a site if it meets all of the following three criteria: significantly occupied ($p \leq 0.05$); significantly enriched over input ($p \leq 0.05$); and significantly enriched over mock IP ($p \leq 0.05$).

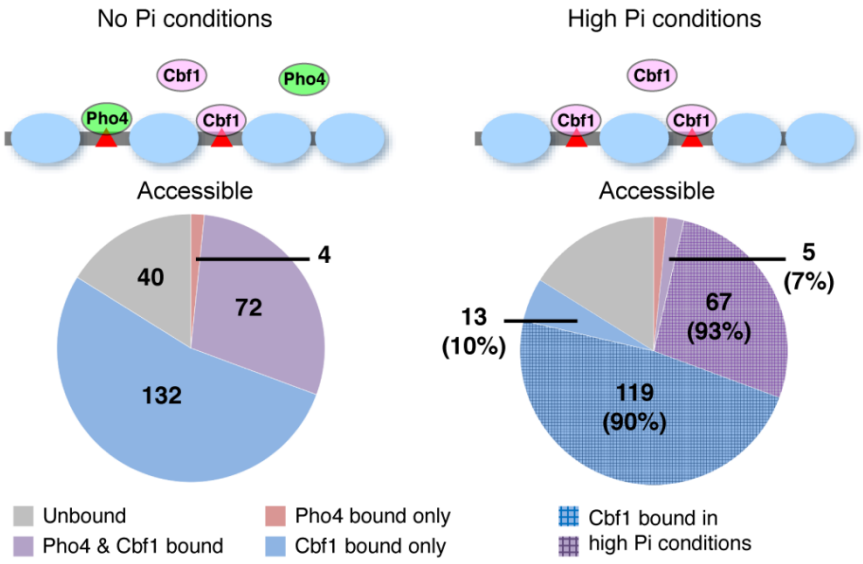


Figure 16. Cbf1 binding at accessible sites in no and high Pi conditions.

On the left, the pie chart shows Cbf1 and Pho4 binding in no Pi conditions. On the right, the hatched area in the pie chart indicates the accessible sites that are bound by Cbf1 in high Pi conditions. A transcription factor is considered bound to a site if it meets all of the following three criteria: significantly occupied ($p \leq 0.05$); significantly enriched over input ($p \leq 0.05$); and significantly enriched over mock IP ($p \leq 0.05$). Cbf1 occupies similar sites in high Pi and no Pi conditions.

To explore the mechanisms by which Pho4 and Cbf1 distinguish these two classes of sites, I examined two possibilities: first, the intrinsic binding affinity to different 'CACGTG' sequences may dictate the differential binding between Pho4 and Cbf1; second, the presence of additional sequence elements may correlate with either class of the sites, which could be an indication for the involvement of other DNA binding proteins.

Pho4 and Cbf1 have different preferences for bases flanking the 'CACGTG' binding site *in vitro* (Maerkl and Quake, 2007); these flanking sequence may explain the differences in Cbf1 and Pho4 occupancy *in vivo*. In accord with the observed *in vitro* sequence preferences, I find that the accessible sites with less Pho4 binding have a single 5' 'T' base flanking the 'CACGTG' (Figure 17), suggesting that Cbf1 can compete most effectively at these sites. The other flanking bases cannot successfully distinguish either the sites preferably bound by Pho4 or the sites favored by Cbf1 (Figure 18). Moreover, the ability of this single 5' 'T' base to differentiate Pho4 preferred binding sites extends beyond the accessible binding sites (Figure 19), suggesting that Cbf1 competes globally with Pho4 for binding. To examine the second possibility, I used MEME to search for sequence motifs that were enriched in the set of sites bound only by Cbf1 or the sites bound by both factors. None of the resulting motifs passed the statistical threshold of enrichment. Thus, the binding pattern of Pho4 and Cbf1 is not likely explained by an interaction with another sequence-specific binding protein. Therefore, the single 'T' flanking base at the 5' end of the 'CACGTG' sequence motif explains the site selection between Pho4 and Cbf1, when Pho4 is nuclear localized due to phosphate starvation signaling.

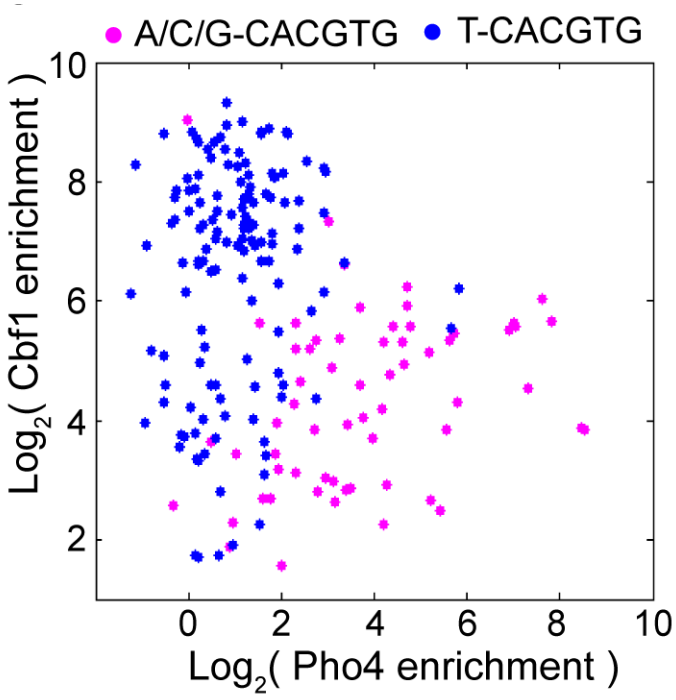


Figure 17. Differential enrichment of Pho4 and Cbf1 binding at accessible high affinity sites with a 'T' flanking sequence.

Scatter plot displays Pho4 and Cbf1 binding enrichment in no Pi conditions for accessible sites that are bound by at least one of the two transcription factors.

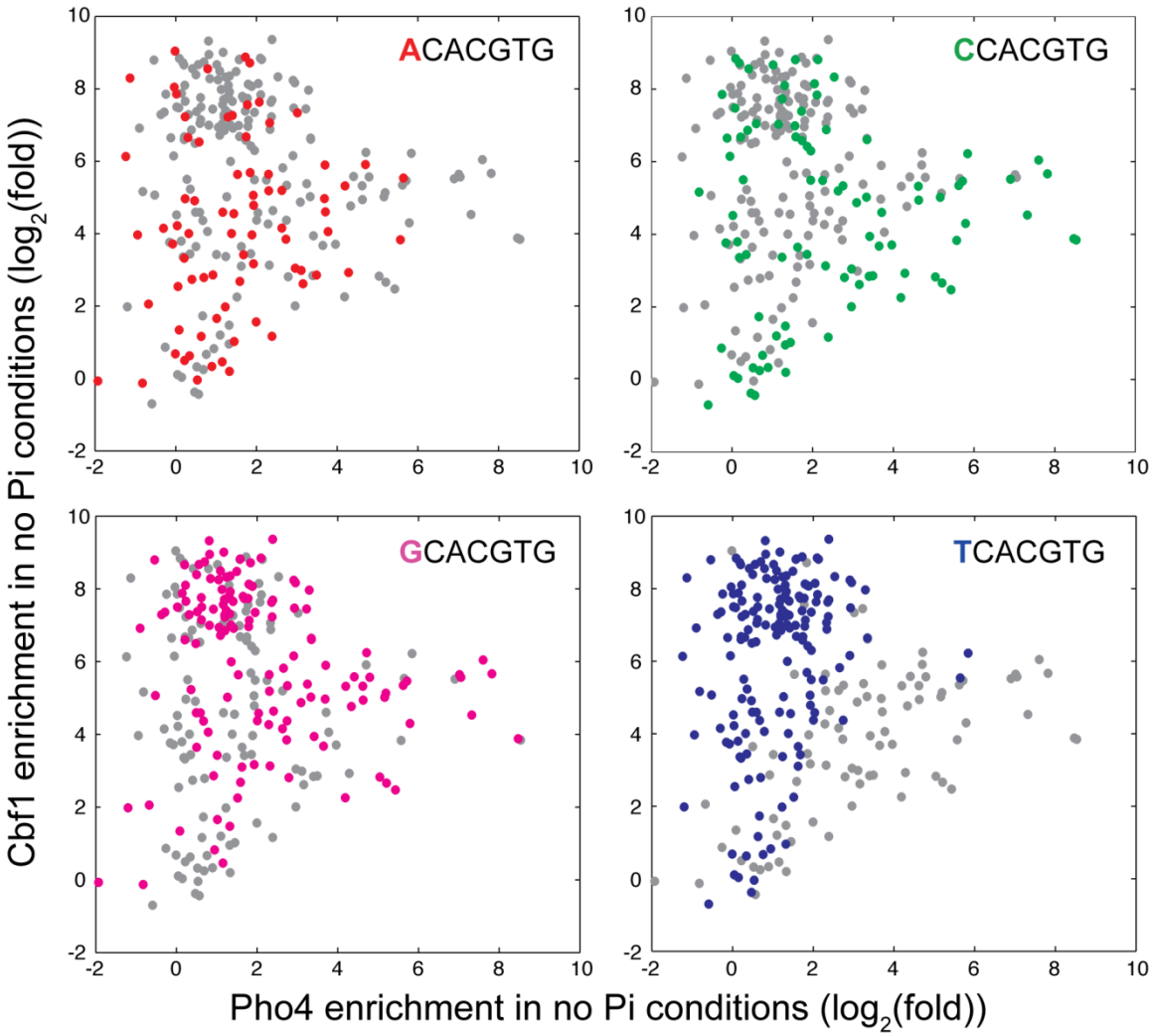


Figure 18. Enrichment of Pho4 and Cbf1 binding at accessible high affinity sites, sorted by flanking sequences.

Scatter plot displays Pho4 and Cbf1 binding enrichment relative to corresponding inputs in no Pi conditions for all accessible high affinity sites.

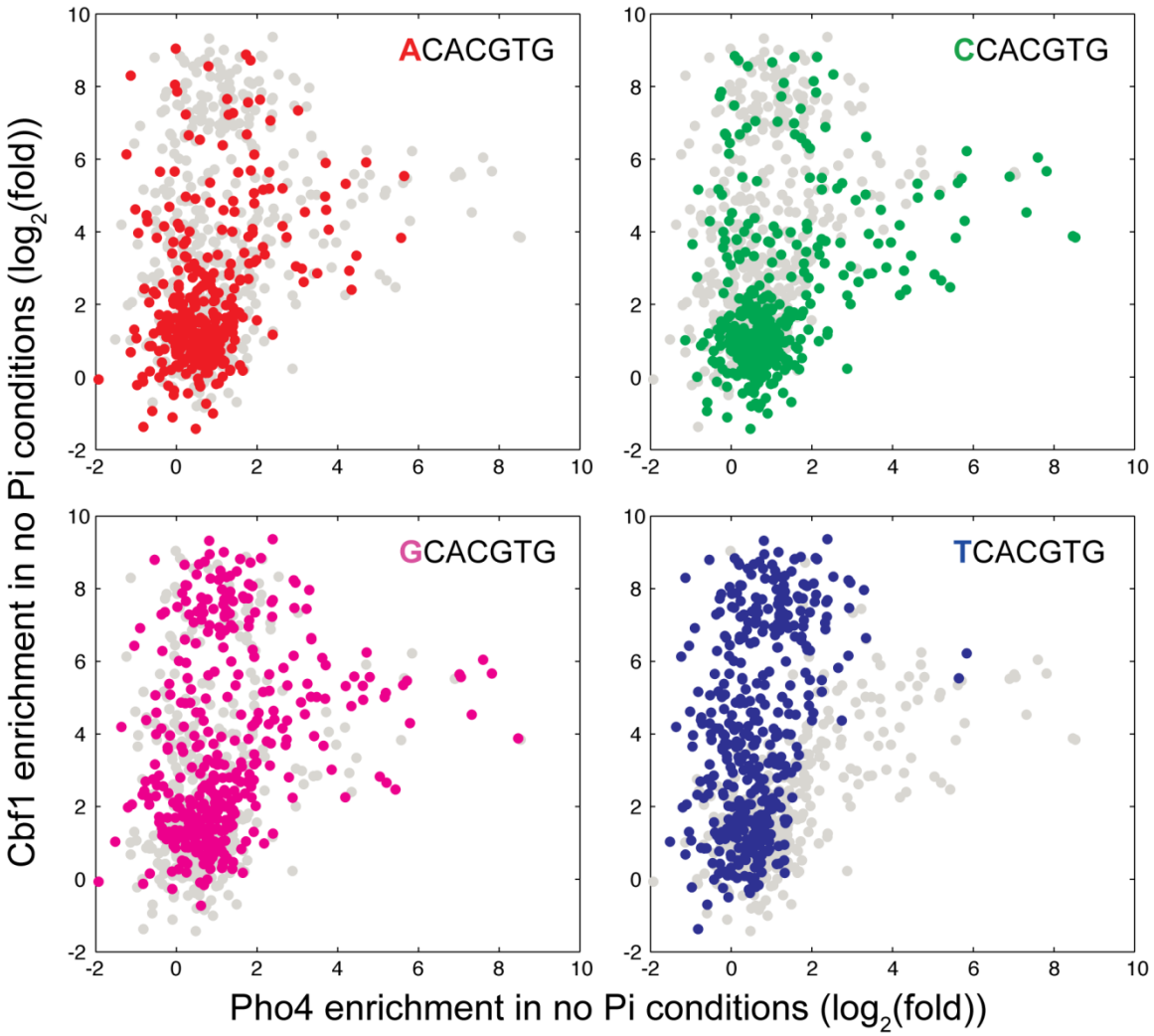


Figure 19. Enrichment of Pho4 and Cbf1 binding at all high affinity sites, sorted by flanking sequences.

Scatter plot displays Pho4 and Cbf1 binding enrichment relative to corresponding inputs in no Pi conditions for all high affinity binding sites.

To demonstrate that Cbf1 competition indeed determines the binding of Pho4 at the accessible sites, I examined the Pho4 binding pattern in strains deleted of *CBF1*. Since deletion

of Cbf1 causes a growth defect and impaired activation of the PHO pathway in no Pi conditions (Figure 20), I performed the experiments using a strain lacking the cyclin Pho80 grown in high Pi medium (O'Neill et al., 1996), in which the PHO pathway is constitutively activated. The binding occupancy of Pho4 in no Pi conditions is almost the same as its binding in *pho80Δ* strains in high Pi conditions (Figure 21, comparing cyan and blue curve), suggesting that the strain with constitutive PHO activation is a good approximation for the wild type strains under phosphate starvation conditions. As shown in Figure 21, comparing Pho4 binding in the *pho80Δ* strain and in the *pho80Δcbf1Δ* strain, the occupancy of pho4 increased dramatically at the sites that are only bound by Cbf1, demonstrating that Cbf1 indeed competes with Pho4 for its binding at these 'CACGTG' sites.

To make sure the change in Pho4 occupancy in the *pho80Δcbf1Δ* strain results exclusively from the lack of Cbf1 competition, I compared Pho4 binding in the *pho80Δ* and *pho80Δcbf1Δ* strains, at the 'TCACGTG' sites. The sites showing an increase in Pho4 binding in the *pho80cbf1Δ* strains are almost exclusively the sites bound by Cbf1 in wild type (compare the blue dots (Cbf1-unbound) against the red dots (Cbf1-bound), Figure 22). The occupancy of Pho4 at the sites that are not bound by Cbf1 remains the same between the *cbf1Δ* and *CBF1* strains (Figure 23). These results demonstrated a correlation between Cbf1 competition and the lack of Pho4 binding at these sites.

It is worth noting that not all of the Cbf1-bound sites gained Pho4 binding in the *pho80Δcbf1Δ* strain - the increase in Pho4 occupancy was dependent on the accessibility of the Cbf1-bound sites. For example, of the 118 Cbf1-bound 'TCACGTG' high affinity sites (Figure 22) showing a 1.6-fold increase in Pho4 binding occupancy, 83 are accessible binding sites for Pho4. Similarly, most of the accessible binding sites that are only occupied by Cbf1 (83 of 119)

showed increase an in Pho4 binding. These data suggest that the influence of Cbf1 competition on the binding of Pho4 is dependent on the binding site accessibility.

Overall, our results demonstrated that the competition from Cbf1 determines the binding occupancy of Pho4 at the nucleosome-free sites; this competition is mediated by their intrinsic preference for specific sequences flanking the core 'CACGTG' binding motif.

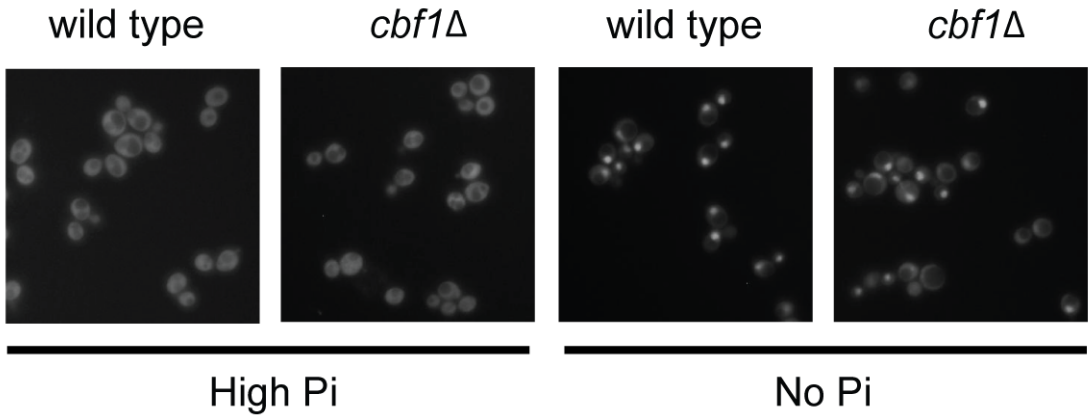


Figure 20. Pho4-GFP localization in wild-type and *cbf1Δ* strains in high Pi and no Pi conditions, respectively.

No difference in Pho4 nuclear localization is observed between wild-type and *cbf1Δ* strains in high Pi conditions. In no Pi conditions, Pho4 shows cell-to-cell variability in localization and incomplete nuclear accumulation in the *cbf1Δ* strain, perhaps as a result of a growth defect.

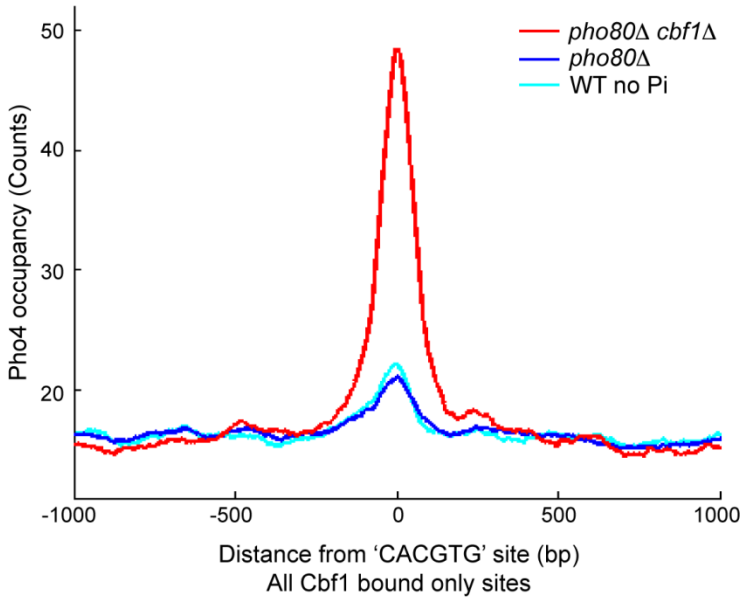


Figure 21. Competition from Cbf1 represses Pho4 binding at the Cbf1-preferred 'CACGTG' sites.

The plots show the average of Pho4 binding occupancy at sites that are only bound by Cbf1 in the wild type. The Pho4 binding occupancy in the *pho80Δ* strain mimics the binding occupancy in the wild type under phosphate starvation conditions. The Pho4 binding occupancy in different strains is normalized by sequencing depth.

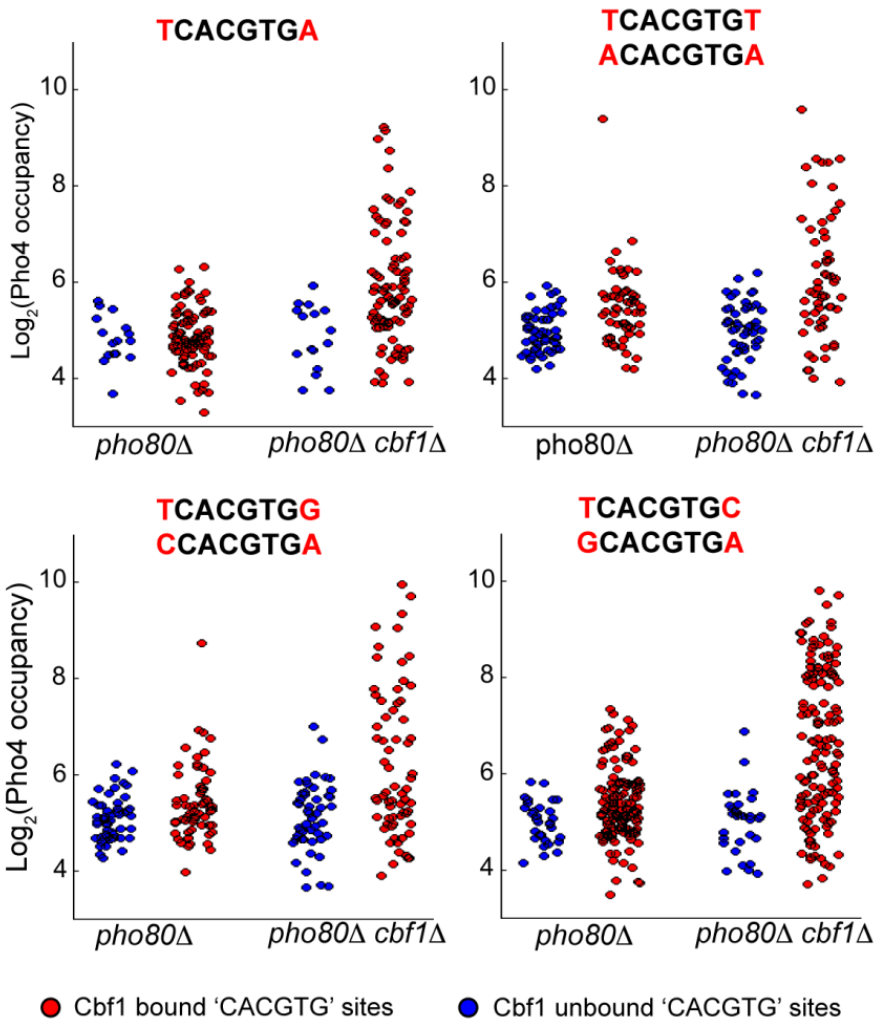


Figure 22. Competition from Cbf1 represses Pho4 binding at Cbf1-preferred 'CACGTG' sites.

Scatter plots show Pho4 binding occupancy at all 'TCACGTG' sites, both accessible and those occluded by nucleosomes. Labeled 8-mer sequences indicate DNA motifs with the same binding preference of Pho4.

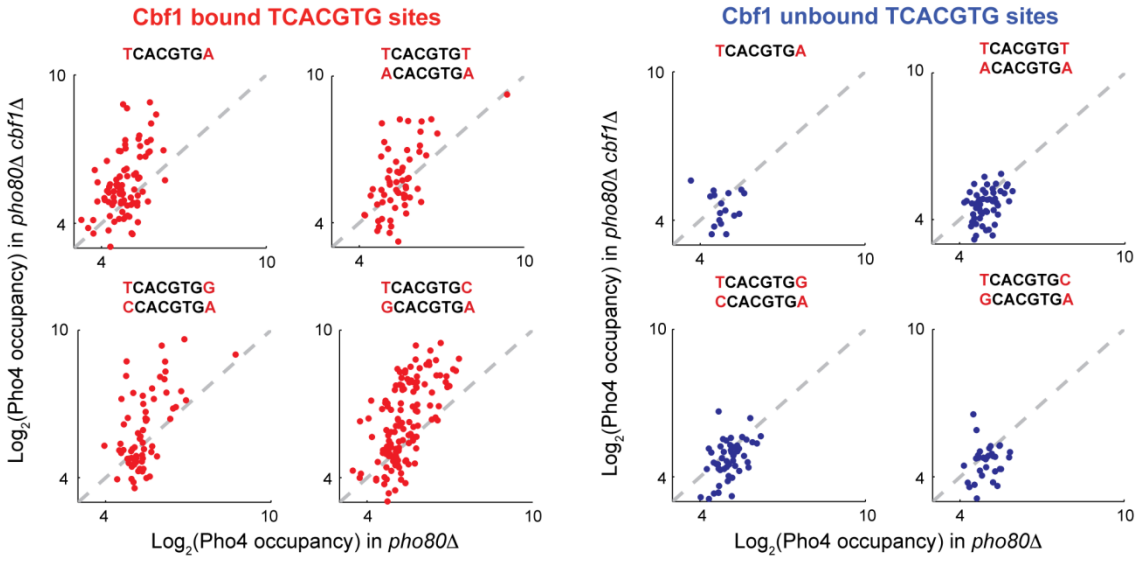


Figure 23. Correlation plots of Pho4 binding occupancy at 'TCACGTG' sites in the *pho80Δ* and *pho80Δ cbf1Δ* strains.

Red dots represent Cbf1-bound 'TCACGTG' sites and blue dots represent Cbf1-unbound 'TCACGTG' sites. The Pho4 binding occupancy in the *pho80Δ* and *pho80Δ cbf1Δ* strains were normalized by the total number of reads.

In summary, I proposed the model below to explain how Pho4 identifies its *in vivo* binding sites among all of its high affinity binding sites in the genome. In high Pi conditions (Figure, left), Pho4 is imported into the nucleus, phosphorylated by cyclin and cyclin dependent kinase complex Pho80-Pho85, and then actively exported into the cytoplasm; thus, the Pho4 nuclear concentration is low (Komeili and O'Shea, 1999; O'Neill et al., 1996). Most of the accessible high affinity binding sites are occupied by Cbf1, which resides in the nucleus constitutively (Huh et al., 2003). In no Pi conditions (Figure, upper right), the kinase activity of Pho80-Pho85 complex is inhibited and Pho4 is no longer exported from the nucleus. Consequently, the nuclear concentration of Pho4 increases, allowing Pho4 to compete effectively for binding at sites that are weakly bound by Cbf1 (Figure, lower right); Cbf1 is bound to the consensus sites with a 5' flanking 'T' with high enough affinity to prevent Pho4 binding.

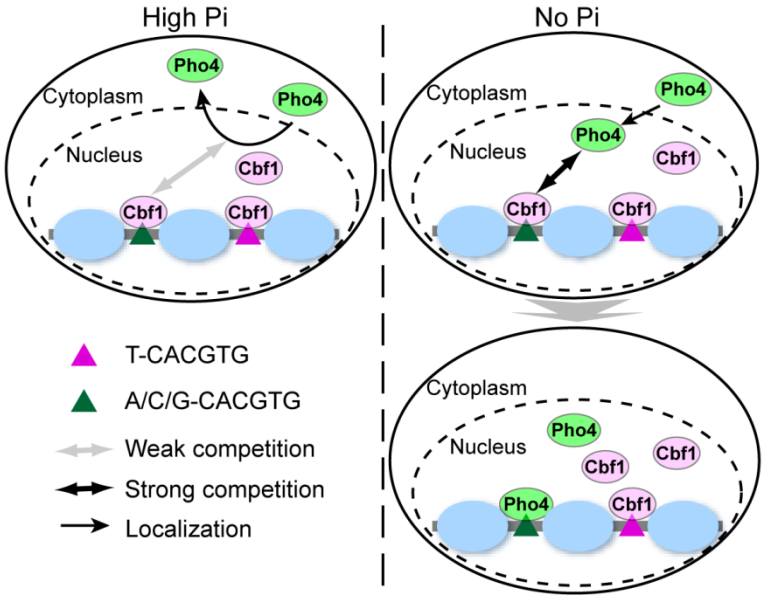


Figure 24. Schematic depicting a model for Pho4 binding to its high affinity sites.

2.6. Competitive binding of Cbf1 influences the activation threshold and specificity of the PHO regulon

What is the physiological role of the interplay between Pho4 and Cbf1 at consensus sites? One idea is that Cbf1 is required in high Pi conditions to keep nucleosomes properly positioned, and therefore keep consensus binding sites nucleosome free – I thus named it "Cbf1 priming model" (Figure 25). This model predicts that the accessibility of the binding site is dependent on the presence of Cbf1, originated from the precedent evidence that Cbf1 positions nucleosomes in the promoters of some sulfur metabolism genes (Kent et al., 2004; Kent et al., 1994). However, when I analyzed the nucleosome occupancy at Pho4 binding sites in the *cbf1Δ* strain in high Pi conditions, I observed no change in nucleosome occupancy at Pho4 binding sites, especially the ones at the promoter of Pho4-regulated genes ($r = 0.943$ between *cbf1Δ* strain and wild-type, $r = 0.953$ between wild-type replicates; Figure 26). In contrast, in the *cbf1Δ* strain nucleosome occupancy increased and nucleosome position shifted at binding sites in the regulatory regions of sulfur metabolism genes (Lee et al., 2010) (Figure 27, Figure 28). Thus, I concluded that Cbf1 does not significantly contribute to the accessibility of the Pho4 binding sites at the promoters of PHO genes.

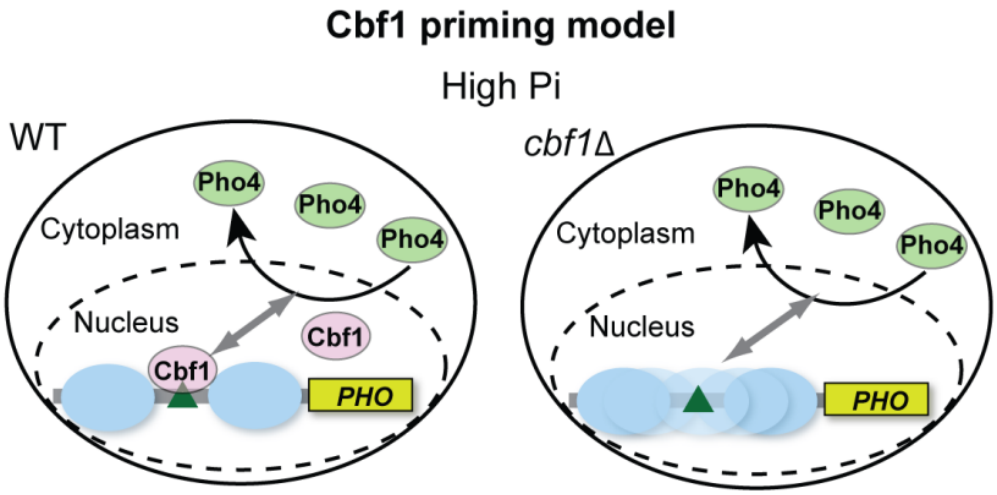


Figure 25. Schematic illustration of the "Cbf1 priming model" in high Pi conditions.

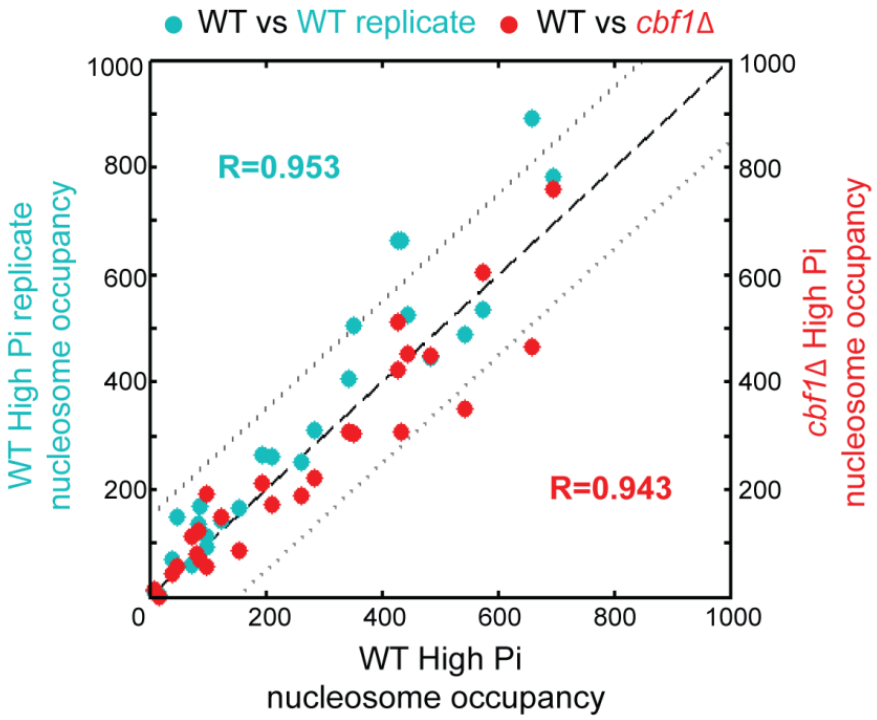


Figure 26. Scatter plots show nucleosome occupancy of wild-type and *cbf1Δ* strains in high Pi conditions at regulated Pho4 binding sites.

Nucleosome occupancy is shown in normalized counts and each colored dot represents a binding site. 'R' denotes the linear correlation coefficient of the dots in the same color. Black dashed lines define the diagonal and gray dashed lines indicate the threshold for nucleosome occupancy of 150 reads from the diagonal. Nucleosome occupancy is unaffected by Cbf1 deletion at regulated Pho4 binding sites.

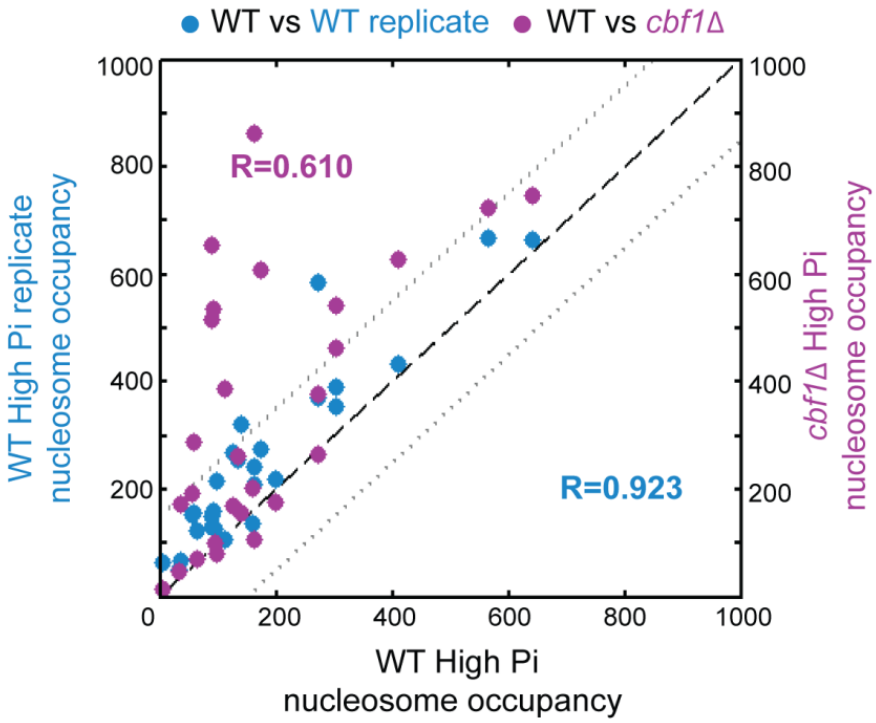


Figure 27. Scatter plots show nucleosome occupancy of wild-type and *cbf1Δ* strains in high Pi conditions at *MET* genes (Lee et al., 2010).

Nucleosome occupancy is shown in normalized counts and each colored dot represents a binding site. 'R' denotes the linear correlation coefficient of the dots in the same color. Black dashed lines define the diagonal and gray dashed lines indicate the threshold for nucleosome occupancy of 150 reads from the diagonal. Nucleosome occupancy at Cbf1 binding sites of *MET* genes (Lee et al., 2010) shows a significant increase in *cbf1Δ* strains in contrast to the nucleosome occupancy at Cbf1 binding sites at *PHO* genes.

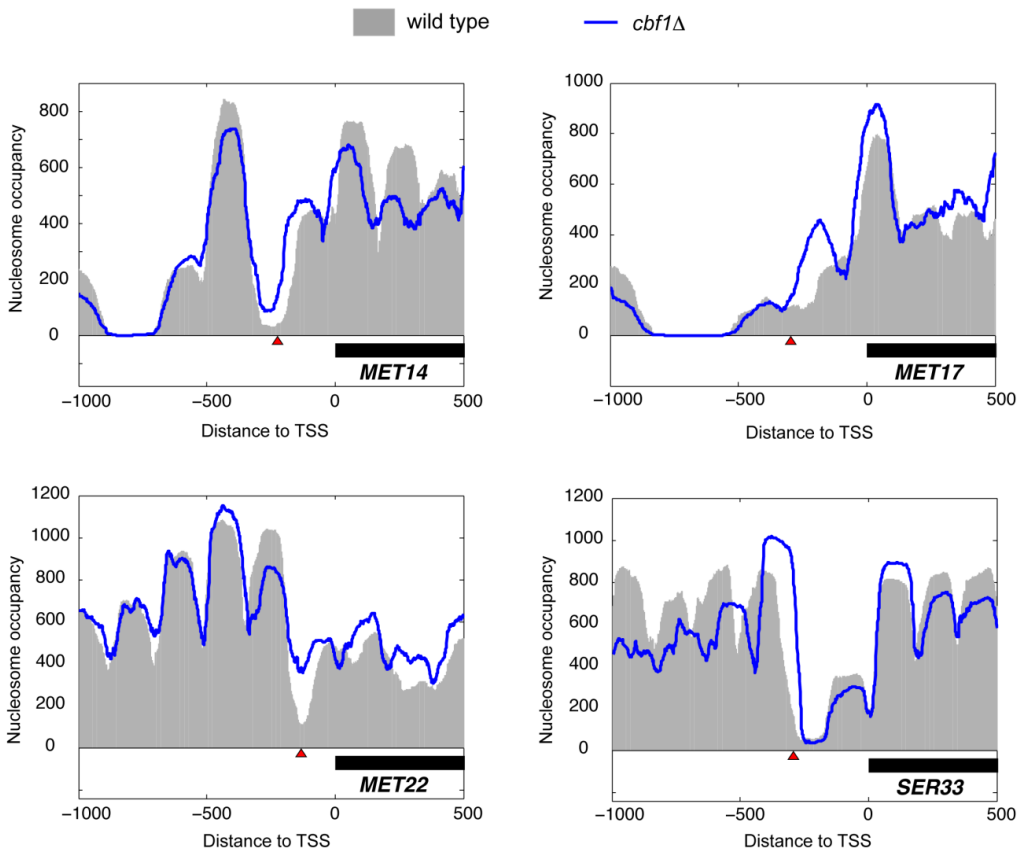


Figure 28. Nucleosome occupancy maps showing changes of nucleosome positions at the promoters of *MET* genes in the *cbf1Δ* strain.

Nucleosome occupancy maps in the wild type strain and the *cbf1Δ* strain were normalized to the same total number of reads. Red triangles indicate the position of a Cbf1 binding site and black boxes indicate the position of the labeled transcript. Shift of nucleosome position is observed at the promoters of *MET14*, *MET17*, *MET22* and *SER33* as the nucleosomes flanking the Cbf1 binding sites move to partially cover the binding sites in the *cbf1Δ* strain.

An alternative model is that competition from Cbf1 raises the binding threshold for Pho4 through sequestering the accessible binding sites. I therefore termed it "Cbf1 blocking model" (Figure 29). This model predicts two possible roles of which Cbf1 competition regulates the transcriptional control by Pho4: in high phosphate conditions, it may prevent spurious activation of phosphate-responsive genes induced by residual nuclear Pho4, and during phosphate starvation conditions, it may prevent Pho4 activating other 'CACGTG' containing genes.

To test these hypotheses, I first examined the role of Cbf1 competition in regulating expression of the PHO genes in high phosphate conditions. In the absence of Cbf1, most Pho4-regulated genes (characterized in Chapter II, section 2.6) should significantly increased expression (20 of 28, $p \leq 0.05$; Figure 30) – conditions where Pho4 is less active and localized primarily to the cytoplasm (Komeili and O'Shea, 1999). This aberrant expression is Pho4-dependent (compare Figure 30, columns 2 and 3), as deletion of Pho4 abolished the aberrant expression of the *PHO* genes in *cbf1Δ* background. In addition, this Pho4 dependent activation of PHO genes was not a result of defects in the upstream signaling of Pho4, because the nuclear localization of Pho4 in the *cbf1Δ* strain is the same as in the wild type strain (Figure 20). Deletion of Rtg3 and Tye7, two other members of the bHLH family that bind the 'CACGTG' consensus site (Zhu et al., 2009a) did not result in similar spurious activation (Figure 30, column 4), indicating that the competition from Cbf1 is the dominant influence among the basic Helix-loop-Helix transcription factors. Thus, the competition from Cbf1 prevents the spurious expression of the PHO genes by low level of nuclear Pho4 in high phosphate conditions, where the PHO pathway is supposed to be turned off.

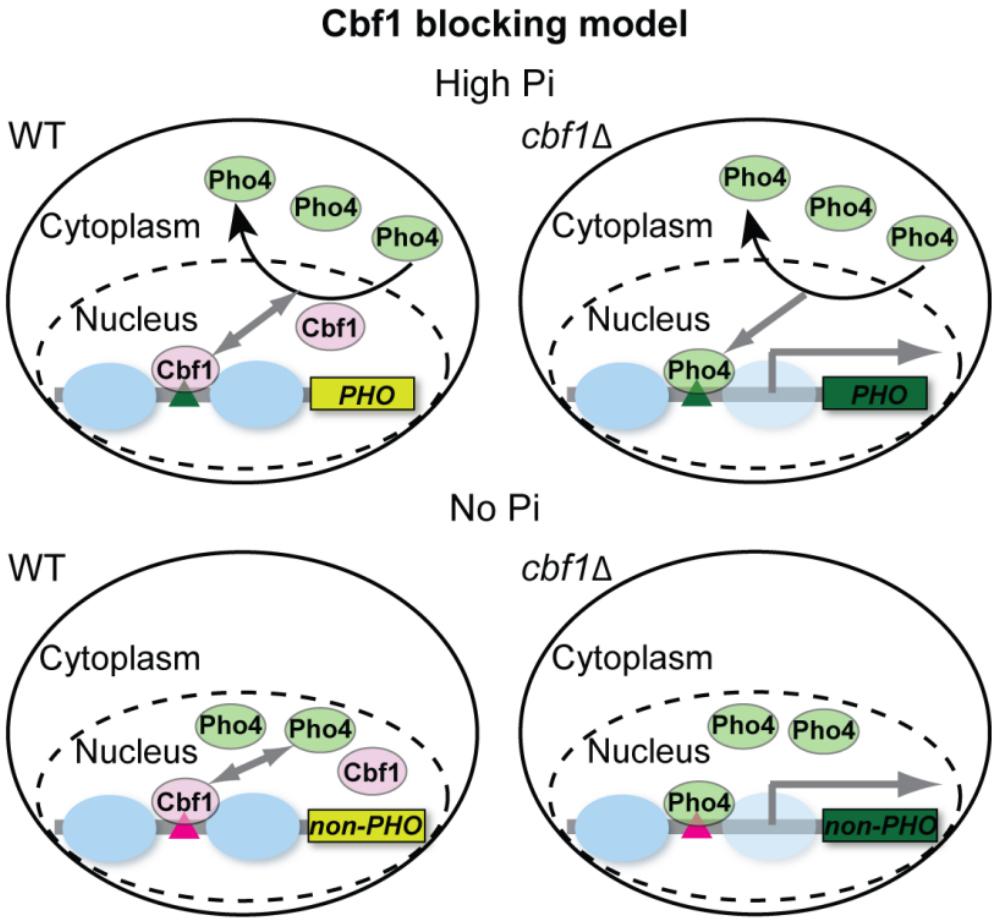


Figure 29. Schematic illustration of the "Cbf1 blocking model" in both high and no Pi conditions.

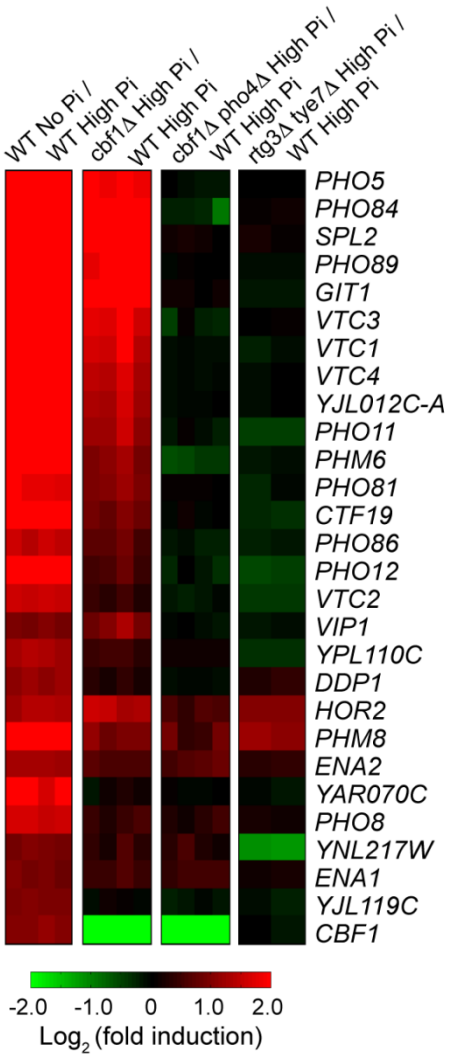


Figure 30. Cbf1 prevents Pho4 spurious transcriptional activation in high Pi conditions.

Heat map showing the fold induction (log_2 scale) of Pho4 regulated genes for wild-type in no Pi conditions (column 1), *cbf1Δ*, *cbf1Δ pho4Δ*, and *rtg3Δ tye7Δ* strains in high Pi conditions (column 2-4), as measured by microarray analysis. Sub-columns indicate biological replicates.

Next, I examined how Cbf1 competition influences Pho4's regulation when PHO pathway is turned on. Since deletion of Cbf1 causes a growth defect and impaired activation of the *PHO* pathway in no Pi conditions (Figure 20), I evaluated the consequences of Cbf1 competition when the *PHO* pathway is fully activated using a strain lacking the cyclin Pho80 grown in high Pi medium (O'Neill et al., 1996). As mentioned earlier, in the *cbf1Δ pho80Δ* strain I observed Pho4 binding to sites that are only bound by Cbf1 in the wild-type strain (Figure 21), particularly at the 'T-CACGTG' motifs (Figure 22, Figure 23). This result led me to explore if the PHO pathway could turn on the expression of genes that are outside of the PHO regulon, especially those associated with Cbf1 binding. I found a class of genes that became regulated by Pho4 in the absence of Cbf1 competition (28 genes, induced more than 2-fold by Pho4 in the *cbf1Δ pho80Δ* strain, and differentially induced more than 1.5-fold comparing to that in the *pho80Δ* strain (Figure 31, left)), 16 of which carry the 'CACGTG' binding sites in their promoters and are clearly bound by Cbf1. The induction of these genes was mostly absent in either wild type under phosphate starvation or the *cbf1Δ* strain in high phosphate conditions (Figure 31, right), indicating that the activation of these genes is a result of lack of Cbf1 competition under the PHO induced conditions. Indeed, Pho4 is bound to and activates transcription of 13 genes that were bound by Cbf1 and not regulated by Pho4 in the wild-type strain (Figure 32).

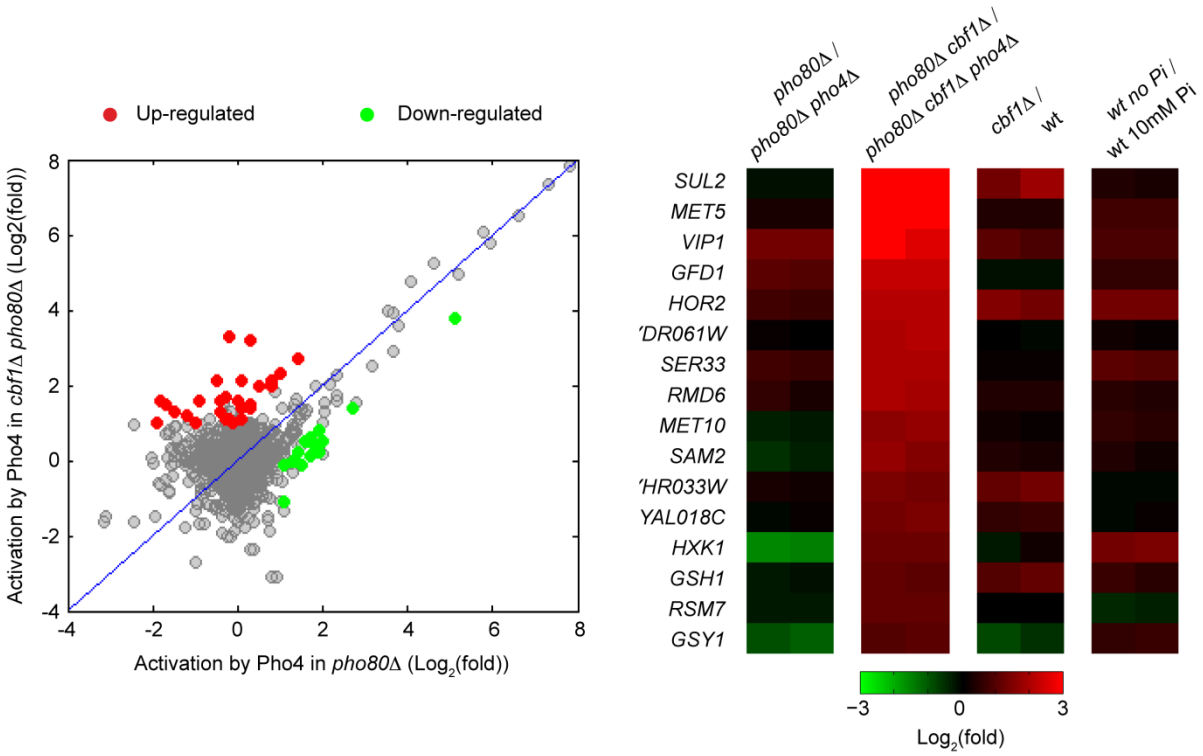


Figure 31. Expression of genes that become Pho4-dependent only in the absence of Cbf1.

On the left, scatter plots show the activation by Pho4 in both the *cbf1Δ pho80Δ* and the *CBF1 pho80Δ* strains. Each dot represents the Pho4-dependent activation of a single gene. The activation is calculated as the logarithm of the expression ratio in the comparison of *pho80Δ* / *pho80Δ pho4Δ*, or *pho80Δ cbf1Δ / pho80Δ cbf1Δ pho4Δ*. The genes with increased induction by Pho4 in the *cbf1Δ* background (28 genes, significantly up-regulated more than 1.5-fold, $p \leq 0.05$; 16 of the 28 genes carry 'CACGTG' sites) are colored in red, and the genes with increased induction by pho4 in the CBF1 background (16 genes) are colored in green. On the right, heat map of the expression difference for the 16 genes that carry the 'CACGTG' binding sites at their promoters.

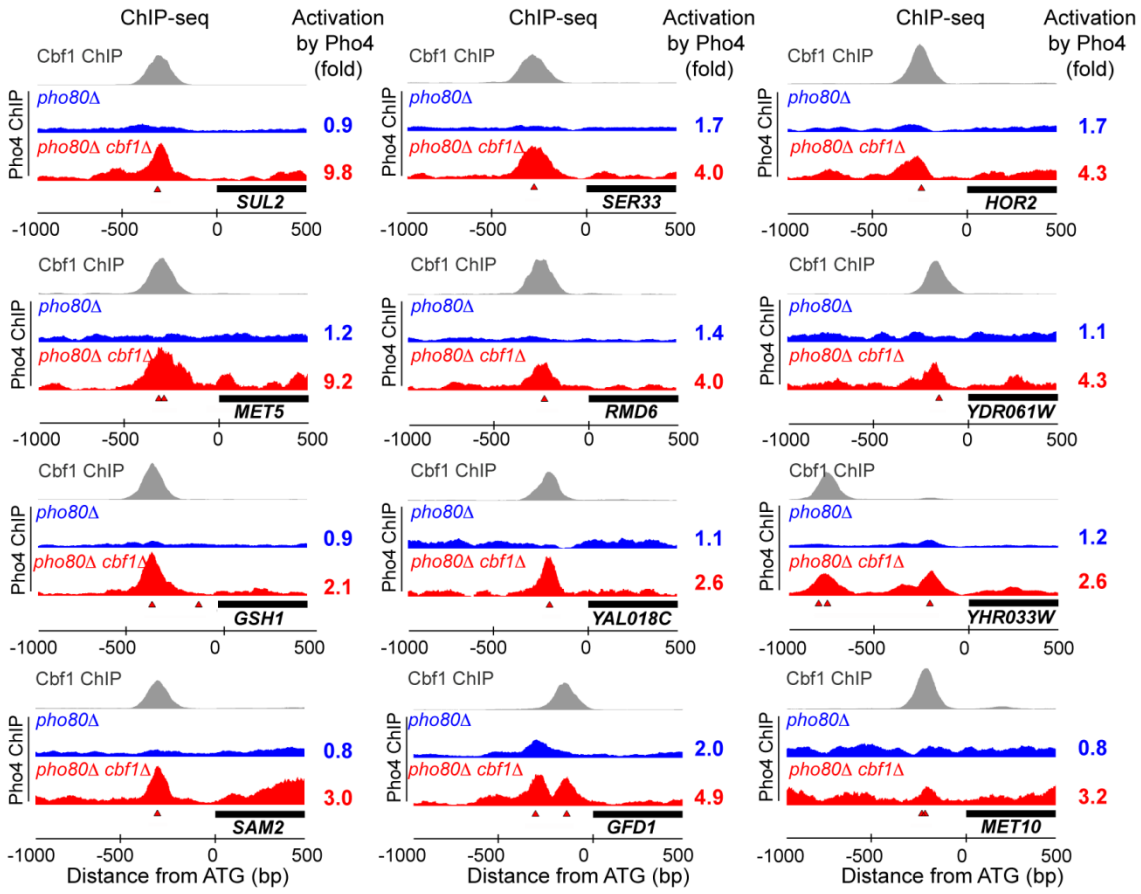


Figure 32. Differential binding and activation by Pho4 in the presence and absence of Cbf1.

Tracks show ChIP-seq results of Cbf1 binding in no Pi conditions (gray), Pho4 binding in a *pho80Δ* strain (constitutively nuclear Pho4, blue) and in a *cbf1Δ pho80Δ* (red) strain. Gene activation by Pho4 in *pho80Δ* (blue) and *cbf1Δ pho80Δ* (red) strains is determined by comparing gene expression in *pho80Δ* and *pho80Δ pho4Δ*, *cbf1Δ pho80Δ* and *cbf1Δ pho80Δ pho4Δ* strains, respectively. Red triangles mark the high affinity 'CACGTG' binding sites.

Our observations support two roles for Pho4-Cbf1 competition: in high Pi conditions Cbf1 prevents spurious activation of the *PHO* genes induced by a low level of nuclear Pho4, ensuring that phosphate-responsive genes are turned off when Pi is available; in no Pi conditions, Cbf1 prevents Pho4 from inappropriately activating genes containing a 'CACGTG' motif that are not part of the phosphate regulon, ensuring that only genes needed for the response to phosphate limitation are turned on.

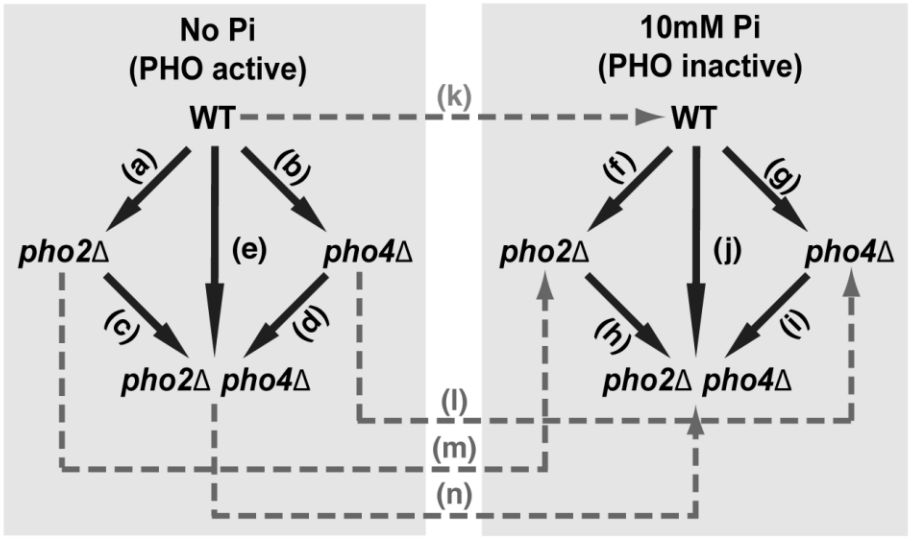
2.7. Dissecting the regulatory interactions in PHO pathway

It is commonly observed that the number identified binding events of transcription factors is far more than the number of genes being regulated by these transcription factors. However, it is yet unclear how transcription factors only select a fraction of their associated genes to regulate, and if/how other co-factors involve in the selection of functional binding events from a mechanistic perspective.

As noted in the previous sections, Pho4 binds to 115 of the high affinity 'CACGTG' sites, scattered in the promoters of over 80 genes, but only ~10-20 genes are regulated by Pho4 in response to Pi starvation as examined by expression microarrays from two previous studies (Ogawa et al., 2000; Springer et al., 2003). Moreover, these regulated genes have different dependence on Pho4 and on Pho2, a homeodomain factor that interacts with Pho4 and regulates the phosphate starvation response (Vogel et al., 1989). For example, several PHO genes show Pho2-dependent activation, including *PHO5*, *PHO89*, *SPL2*, *PHO84* and several other PHO genes (Springer et al., 2003); whereas the induction of *PHO8* in response to Pi starvation does not require the presence of Pho2 (Munsterkotter et al., 2000). Therefore, it is likely that Pho2 plays a pivotal role in activating gene transcription among all of the genes associated with Pho4 binding. To test this hypothesis, I need to examine how Pho4 and Pho2 contribute to the transcriptional activation of the genes responsive to phosphate starvation, and connect their contribution with the binding of these factors.

An epistasis expression analysis approach (mutant cycle analysis) was developed in the lab to analyze the regulatory interactions among multiple transcription factors involved in osmotic stress in yeast (Capaldi et al., 2008). I adapted this approach to dissect and quantify the

contribution of the regulatory interactions between Pho2 and Pho4 to transcriptional activation in response to phosphate starvation (Figure 33). These regulatory interactions can be described in terms of three "expression components": the contribution of Pho2 acting alone (Pho2), the contribution of Pho4 acting alone (Pho4), as well as the contribution of Pho2 and Pho4 acting together (which I refer to as a "cooperative" component). To determine the values of these expression components, I directly compared gene expression between all possible pairs of wild-type, *pho2Δ*, *pho4Δ* and *pho2Δ pho4Δ* strains in no and high Pi conditions, where the measured gene expression difference between each pair of strains consists of a linear sum of the expression components relevant for that pairwise comparison (Figure 33). In no Pi conditions, 60 minutes of Pi starvation was chosen because at this time point the *PHO* genes are sufficiently activated and secondary effects are not significantly observed (data not shown). I calculated the expression components for each gene through regression methods with equations describing the expression components measured in each microarray and the details are elaborated below.



(a) Pho2 + Co + $Pho2_B$ + Co_B
(b) Pho4 + Co + $Pho4_B$ + Co_B
(c) Pho4 + $Pho4_B$
(d) Pho2 + $Pho2_B$
(e) Pho2 + Pho4 + Co + $Pho2_B$ + $Pho4_B$ + Co_B

(k) Pho2 + Pho4 + Co + K
(l) Pho2 + K
(m) Pho4 + K
(n) K

(f) $Pho2_B$ + Co_B
(g) $Pho4_B$ + Co_B
(h) $Pho4_B$
(i) $Pho2_B$
(j) $Pho2_B$ + $Pho4_B$ + Co_B

Figure 33. Design of epistasis expression analysis for phosphate starvation response.

14 comparisons (a - n) between wild-type, *pho2Δ, pho4Δ*, and *pho2Δ pho4Δ* strains in both high and no Pi conditions comprise mutant cycle analysis. Arrowheads indicate the samples labeled with Cy5 dye and arrow tails indicate the samples labeled with Cy3 dye in each microarray hybridization. Each arrow is labeled with a letter that can be decomposed into expression components that contribute to gene induction in each comparison. The components are listed below the mutant cycle diagram with the indicated letter. K presents the induction in no Pi conditions that is independent of Pho2 and Pho4. Pho2, Pho4 and Co present activation from Pho2 alone, Pho4 alone and activation dependent on both Pho2 and Pho4 in no Pi conditions. $Pho2_B$, $Pho4_B$ and Co_B present expression regulated by Pho2 alone, Pho4 alone and both Pho2 and Pho4 in high Pi condition.

Each experimental expression measurement can be described as the sum of the following components: Pho2 (the influence of Pho2 alone in no Pi conditions), Pho4 (the influence of Pho4 alone in no Pi conditions), Co (the effect of the interaction between Pho2 and Pho4 in no Pi conditions), $Pho2_B$ (the influence of Pho2 alone in high Pi conditions), $Pho4_B$ (the influence of Pho4 alone in high Pi conditions), Co_B (the effect of the interaction between Pho2 and Pho4 in high Pi conditions), and *k* (the influence independent of either Pho2 or Pho4 in no Pi conditions). The measured gene expression difference between each pairwise comparison can be formulated as below:

	Microarray comparison	**Expression components**
A	wt no Pi vs *pho2Δ* no Pi	$Pho2 + Co + Pho2_B + Co_B$
B	wt no Pi vs *pho4Δ* no Pi	$Pho4 + Co + Pho4_B + Co_B$
C	*pho2Δ* no Pi vs *pho2Δ pho4Δ* no Pi	$Pho4 + Pho4_B$
D	*pho4Δ* no Pi vs *pho2Δ pho4Δ* no Pi	$Pho2 + Pho2_B$
E	wt no Pi vs *pho2Δ pho4Δ* no Pi	$Pho2 + Pho4 + Co + Pho2_B + Pho4_B + Co_B$
F	wt high Pi vs *pho2Δ* high Pi	$Pho2_B + Co_B$
G	wt high Pi vs *pho4Δ* high Pi	$Pho4_B + Co_B$
H	*Pho2Δ* high Pi vs *pho2Δ pho4Δ* high Pi	$Pho4_B$
I	*Pho4Δ* high Pi vs *pho2Δ pho4Δ* high Pi	$Pho2_B$
J	wt vs high Pi vs *pho2Δ pho4Δ* high Pi	$Pho2_B + Pho4_B + Co_B$
K	wt no Pi vs wt high Pi	$Pho2 + Pho4 + Co + k$
L	*pho4Δ* no Pi vs *pho4Δ* high Pi	$Pho2 + k$
M	*pho2Δ* no Pi vs *pho2Δ* high Pi	$Pho4 + k$
N	*pho2Δpho4Δ* no Pi vs *pho2Δpho4Δ* high Pi	k

The above formula can be then transformed into matrix multiplication in the form of $\boldsymbol{Y} = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, as shown below:

$$\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \\ \boldsymbol{C} \\ \boldsymbol{D} \\ \boldsymbol{E} \\ \boldsymbol{F} \\ \boldsymbol{G} \\ \boldsymbol{H} \\ \boldsymbol{I} \\ \boldsymbol{J} \\ \boldsymbol{K} \\ \boldsymbol{L} \\ \boldsymbol{M} \\ \boldsymbol{N} \end{pmatrix} = \begin{pmatrix} 1&0&1&1&0&1&0 \\ 0&1&1&0&1&1&0 \\ 0&1&0&0&1&0&0 \\ 1&0&0&1&0&0&0 \\ 1&1&1&1&1&1&0 \\ 0&0&0&1&0&1&0 \\ 0&0&0&0&1&1&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&0&1&0&0 \\ 0&0&0&1&1&1&0 \\ 1&1&1&0&0&0&1 \\ 1&0&0&0&0&0&1 \\ 0&1&0&0&0&0&1 \\ 0&0&0&0&0&0&1 \end{pmatrix} \times \begin{pmatrix} \\ \text{Pho2} \\ \text{Pho4} \\ \text{Co} \\ \text{Pho2}_{\text{B}} \\ \text{Pho4}_{\text{B}} \\ \text{Co}_{\text{B}} \\ k \\ \\ \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon_A} \\ \boldsymbol{\varepsilon_B} \\ \boldsymbol{\varepsilon_C} \\ \boldsymbol{\varepsilon_D} \\ \boldsymbol{\varepsilon_E} \\ \boldsymbol{\varepsilon_F} \\ \boldsymbol{\varepsilon_G} \\ \boldsymbol{\varepsilon_H} \\ \boldsymbol{\varepsilon_I} \\ \boldsymbol{\varepsilon_J} \\ \boldsymbol{\varepsilon_K} \\ \boldsymbol{\varepsilon_L} \\ \boldsymbol{\varepsilon_M} \\ \boldsymbol{\varepsilon_N} \end{pmatrix}$$

For each gene, $\boldsymbol{Y}$ represents the expression change measured in all comparisons, $\boldsymbol{X}$ denotes the design matrix, $\boldsymbol{\beta}$ represents the contribution of seven components, and $\boldsymbol{\varepsilon}$ represents noise. Linear regression analysis can be performed (Capaldi et al., 2008) to infer the component value $\boldsymbol{\beta}$ from experimental measured $\boldsymbol{Y}$, in the condition of minimizing the noise (error). Three biological replicates (42 measurements) were analyzed together by replicating the design matrix and the statistical significance was calculated as described (Capaldi et al., 2008). Examples of the expression measurement and the calculated expression components are shown in Figure 34. *HIS4*, a gene whose basal expression is known to be regulated by Pho2 and Bas1 (a transcription factor involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways (Daignan-Fornier and Fink, 1992; Tice-Baldwin et al., 1989)), has a significant basal Pho2 expression component ($Pho2_B$) but was not regulated by either Pho2 or Pho4 in response to phosphate starvation. A similar result was also observed for *ADE17*, another gene in the purine biosynthesis pathway. Intriguingly, *PHO84*, a PHO gene rapidly induced in phosphate starvation conditions, is regulated by the interaction of Pho2 and Pho4 for basal expression as well as in response to Pi starvation. This surprising result suggests that the low level of nuclear Pho4 in the high phosphate conditions is able to regulate the physiological transcription of genes with a low activation threshold (Lam et al., 2008), supporting the previous result that reducing the activation threshold by removing the competition from Cbf1 can induce PHO gene transcription even in phosphate replete conditions. Overall, the epistasis expression analysis was able to dissection the gene regulatory interaction between Pho2 and Pho4 at specific conditions, providing us with unprecedented information about the transcriptional control of the PHO pathway.

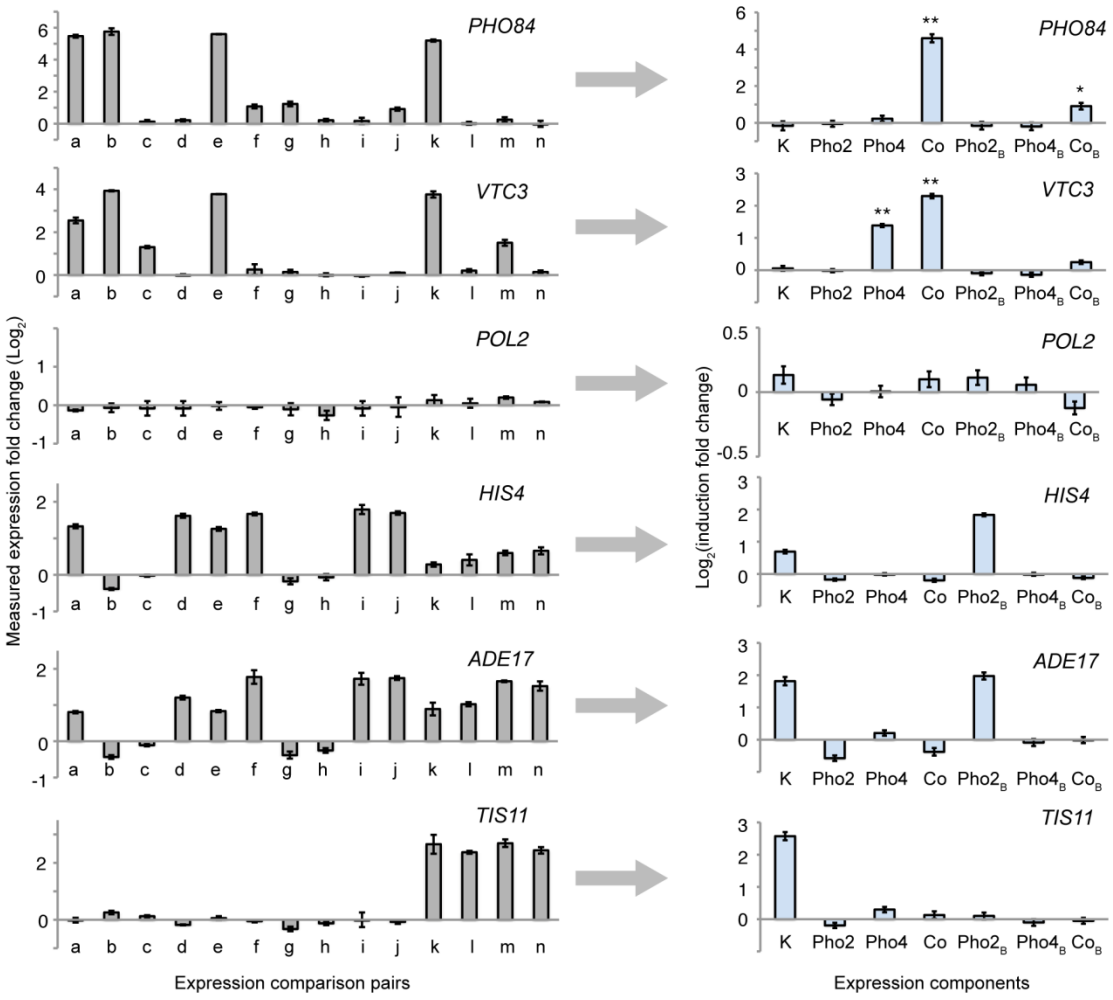


Figure 34. Examples of the expression measurements and expression components inferred from the epistasis expression analysis.

Left, gray bars indicate measured expression fold change for *PHO84*, *VTC3, POL2, HIS4, ADE17 and TIS11* in 14 microarray comparisons listed in Figure 33. Error bard indicates the standard deviation from three independent measurements. Right, light blue bars show the expression components inferred from the measured expression fold change. Stars on top of the bars indicate the significance of the inferred expression component against a null hypothesis that the expression component is less than 1.4-fold. Single star $p \leq 0.05$, double stars $p \leq 0.01$.

80 genes were selected as Pi starvation responsive genes by directly comparing the expression of wild-type in no and high Pi conditions (activated more than 1.8-fold and significantly induced with a null hypothesis of < 1.5 fold, $p \leq 0.01$). For each of these genes, the above linear regression was performed and genes with significant Pho2, Pho4, or Co components (significantly induced with a null hypothesis of < 1.4 fold, $p \leq 0.05$, (Capaldi et al., 2008)) were shown in Figure 35. Full mutant cycle analysis results are listed in Table 4. None of the genes induced after Pi starvation contained a significant Pho2-only component, indicating that Pho4 is the primary activator and Pho2 only functions as an auxiliary factor. This is in fact consistent with the notion that Pho2 is not regulated by environmental phosphate availability (Komeili and O'Shea, 1999; Springer et al., 2003).

I observed only three regulatory interactions operating at phosphate-responsive genes: gene activation by Pho4 acting alone (Pho4 only), genes activated by the interaction between Pho2 and Pho4 (Cooperative, Co), and mixed regulation (genes partially activated by Pho4 alone and partially dependent on the interaction between Pho2 and Pho4, Pho4 + Co) (Figure 35**)**. Notably, almost all Pho4 regulated genes contain a cooperative component except *PHO8*. For the set of gene with mixed regulation between Pho2 and Pho4, the value of cooperative components is much higher than the value of the Pho4 only components. Taken together, these results suggested that the interaction between Pho2 and Pho4 is necessary for the sufficient gene activation.

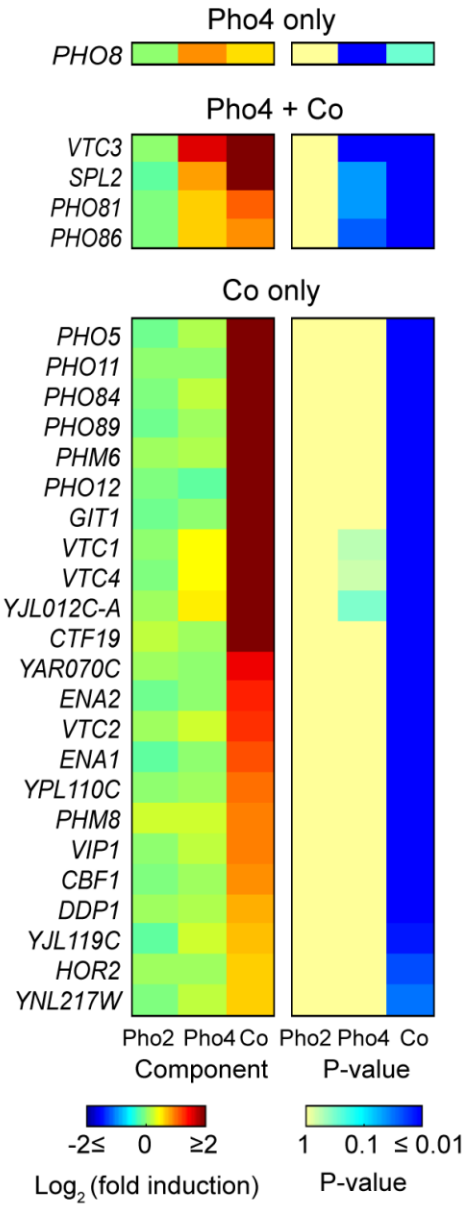


Figure 35. Epistasis expression analysis of phosphate starvation response.

Heat map showing the best fit of expression components (left columns) and their statistical significance (right columns) for genes that regulated by Pho2 and Pho4. Genes are clustered according to the statistical significance ($p \leq 0.05$) of the Pho2, Pho4 and cooperative component (Co).

Table 4. Mutant cycle results of genes induced by phosphate starvation.

ORF	NAME	Expression Components							
		Error	k	Pho2	Pho4	Co	$Pho2_B$	$Pho4_B$	Co_B
YAL005C	SSA1	0.2	0.7	0.2	0.0	0.4	-0.3	0.1	0.1
YAR068W	YAR068W	-0.3	0.5	0.1	-0.2	0.6	0.0	0.3	-0.3
YAR070C	YAR070C	0.2	-0.1	0.1	0.0	1.4	0.0	0.1	0.0
YAR071W	PHO11	-0.3	-0.3	0.1	0.1	5.4	0.2	0.1	0.1
YBR072W	HSP26	0.0	2.1	0.6	0.1	-0.9	0.0	0.2	0.0
YBR093C	PHO5	-0.1	0.1	0.0	0.2	6.4	0.0	0.2	0.1
YBR157C	ICS2	-0.1	1.0	0.1	0.0	0.4	0.2	0.1	-0.1
YBR169C	SSE2	0.1	0.9	0.2	0.1	-0.4	-0.3	0.0	0.3
YBR296C	PHO89	-0.1	0.4	-0.1	0.1	5.1	0.1	0.2	0.0
YCL040W	GLK1	0.0	0.7	0.4	0.2	-0.3	-0.2	0.0	0.1
YCR098C	GIT1	-0.2	0.3	0.0	0.2	3.9	0.1	0.2	-0.1
YDL204W	YDL204W	0.1	1.6	0.2	-0.2	-0.5	0.0	0.4	-0.1
YDR005C	MAF1	-0.2	0.4	0.1	0.0	0.7	0.1	0.2	-0.1
YDR019C	GCV1	0.1	1.5	-0.1	-0.2	0.0	0.3	0.2	-0.1
YDR039C	ENA2	-0.1	0.2	-0.1	0.1	1.3	0.3	0.0	-0.2
YDR040C	ENA1	0.0	0.2	-0.2	0.1	1.1	0.4	0.1	-0.4
YDR270W	CCC2	0.0	1.1	-0.1	0.1	-0.2	0.0	-0.2	0.1
YDR281C	PHM6	-0.3	0.2	0.2	0.3	4.1	0.0	0.0	0.1
YDR481C	PHO8	0.1	0.2	0.1	1.0	0.4	0.1	0.0	0.1
YDR516C	EMI2	0.0	1.0	0.3	0.0	-0.4	-0.2	0.0	0.2
YEL011W	GLC3	0.0	1.0	0.4	0.0	-0.3	-0.2	0.2	0.0
YEL065W	SIT1	0.1	2.1	-0.3	0.1	-0.1	0.2	-0.2	-0.1
YER037W	PHM8	0.2	0.7	0.2	0.2	1.1	0.0	0.0	-0.1
YER062C	HOR2	0.1	0.3	0.1	0.1	0.7	-0.2	0.0	0.2
YER072W	VTC1	-0.2	0.2	0.0	0.5	2.8	0.0	0.0	0.1
YFL004W	VTC2	0.0	-0.1	0.1	0.3	1.3	0.0	-0.1	0.0
YFR053C	HXK1	0.0	0.9	0.4	0.0	-0.4	-0.5	0.2	0.3
YGR233C	PHO81	0.0	0.1	0.0	0.7	1.0	0.0	0.0	0.0
YHL035C	YHL035C	0.1	1.7	0.0	0.0	-0.3	0.1	-0.1	0.1
YHL040C	ARN1	0.0	2.4	0.0	0.0	-0.6	-0.1	0.0	0.1
YHL047C	ARN2	0.0	3.3	0.1	-0.1	-0.9	-0.1	0.1	0.0
YHR136C	SPL2	0.2	0.2	-0.2	1.0	4.2	0.3	0.2	-0.1
YHR138C	YHR138C	0.0	0.8	-0.1	-0.1	0.2	0.0	0.0	-0.1
YHR214W-A	YHR214W-A	-0.3	0.5	0.2	-0.2	0.6	0.0	0.4	-0.3
YHR215W	PHO12	-0.2	-0.3	0.1	-0.1	4.4	0.3	0.2	0.0
YIL074C	SER33	0.1	0.4	0.1	0.0	0.5	0.1	0.0	0.0
YIL169C	YIL169C	-0.2	0.3	0.2	-0.2	0.5	-0.1	0.3	-0.3
YJL012C	VTC4	0.1	0.0	0.0	0.5	2.6	0.0	0.1	0.1
YJL012C-A	YJL012C-A	-0.1	0.0	0.1	0.6	2.5	0.1	0.0	0.1
YJL117W	PHO86	0.0	0.0	0.0	0.6	1.0	0.0	0.0	0.0
YJL119C	YJL119C	0.0	0.4	-0.1	0.4	0.8	0.2	0.2	-0.2
YJR060W	CBF1	0.0	0.1	0.0	0.1	1.0	0.1	0.1	-0.1

Table 4. Mutant cycle results of genes induced by phosphate starvation (Continued).

ORF	NAME	Expression Components							
		Error	k	Pho2	Pho4	Co	$Pho2_B$	$Pho4_B$	Co_B
YKL001C	MET14	-0.1	1.0	0.3	-0.1	-0.3	0.1	0.0	0.0
YKR034W	DAL80	0.0	1.2	0.1	0.0	-0.4	0.4	-0.1	0.3
YKR080W	MTD1	0.0	0.5	0.0	-0.1	0.4	0.2	0.0	0.0
YLL026W	HSP104	0.2	0.6	0.1	0.0	0.1	0.0	0.1	-0.1
YLR109W	AHP1	0.1	0.9	-0.3	-0.2	0.4	-0.2	-0.1	0.2
YLR136C	TIS11	-0.1	2.6	0.0	0.3	0.1	0.0	0.0	0.0
YLR142W	PUT1	0.1	1.2	0.3	0.1	-0.5	0.1	-0.1	0.2
YLR214W	FRE1	0.0	2.2	-0.3	0.0	-0.2	0.2	-0.1	-0.2
YLR303W	MET17	0.1	1.3	0.1	0.0	-0.4	-0.1	-0.1	0.1
YLR327C	YLR327C	0.1	2.5	0.0	-0.4	-0.8	-0.2	0.2	0.0
YLR410W	VIP1	0.1	-0.2	-0.1	0.2	1.0	0.0	0.0	0.0
YLR438W	CAR2	-0.1	0.9	0.2	0.1	-0.5	0.2	0.0	0.2
YML123C	PHO84	-0.6	0.7	0.0	0.6	4.9	0.2	0.0	1.2
YML128C	MSC1	0.1	1.4	0.1	-0.2	-0.2	0.0	0.3	0.0
YMR011W	HXT2	-0.1	1.0	0.0	0.2	0.0	0.4	0.0	-0.2
YMR058W	FET3	0.1	1.8	-0.1	0.2	-0.3	0.2	-0.1	-0.1
YMR105C	PGM2	0.0	0.8	0.4	0.0	-0.3	-0.3	0.0	0.3
YMR173W	DDR48	0.0	0.6	0.0	0.0	0.3	0.1	0.0	0.0
YMR173W-A	YMR173W-A	0.0	0.9	0.0	0.0	0.3	0.2	0.1	0.0
YMR195W	ICY1	0.0	0.6	0.1	0.1	0.3	0.2	0.0	0.1
YMR251W	YMR251W	-0.1	1.4	0.2	0.1	-0.2	0.0	0.0	0.1
YMR251W-A	HOR7	-0.5	1.1	0.1	-0.2	0.6	0.1	0.6	-0.1
YNL217W	YNL217W	0.1	0.1	0.0	0.2	0.7	0.2	-0.1	-0.1
YNR069C	YNR069C	0.1	0.6	0.2	-0.2	0.4	-0.3	0.1	0.1
YOL086C	ADH1	-0.2	0.9	0.6	0.3	-0.7	-0.6	0.0	0.4
YOL155C	YOL155C	-0.4	0.4	0.2	-0.1	0.5	0.0	0.2	-0.3
YOL158C	ENB1	0.0	2.9	0.2	0.1	-0.6	-0.1	-0.1	0.2
YOR163W	DDP1	0.1	0.1	-0.1	0.1	0.8	0.0	0.0	0.0
YOR173W	YOR173W	0.0	1.0	0.3	-0.1	-0.3	-0.2	0.2	0.1
YOR344C	TYE7	0.2	0.9	-0.1	0.0	0.4	-0.1	0.0	0.2
YOR347C	PYK2	-0.2	0.6	0.3	0.0	0.1	-0.1	0.1	0.2
YOR382W	FIT2	-0.1	3.9	0.2	-0.1	-0.8	0.1	0.1	-0.3
YOR383C	FIT3	-0.4	1.7	0.4	0.0	-1.0	0.0	0.1	-0.1
YPL018W	CTF19	-0.1	-0.1	0.2	0.2	2.0	0.0	0.0	0.0
YPL019C	VTC3	0.0	0.1	0.0	1.6	2.4	0.0	-0.1	0.2
YPL054W	LEE1	0.0	0.5	0.4	0.1	0.4	-0.2	0.0	0.3
YPL110C	YPL110C	-0.1	0.2	0.0	0.1	1.1	0.0	0.0	0.0
YPR167C	MET16	0.0	0.6	0.2	0.0	-0.1	0.0	-0.1	0.1

2.8. Cooperative binding between Pho2 and Pho4 determines the functionality of Pho4 binding events

The epistasis expression analysis revealed that both Pho2 and Pho4 are required to sufficiently induce the transcription of PHO genes. In addition, previous studies have shown that Pho2 and Pho4 interact and bind cooperatively to the *PHO5* promoter (Barbaric et al., 1998; Barbaric et al., 1996; Vogel et al., 1989). I thus hypothesized that this cooperative binding might be correlated with the functionality of Pho4 binding events – the ability to trigger gene activation. I observed a strong correlation between Pho2 and Pho4 binding occupancy in no Pi conditions ($r = 0.926$, $p < 10^{-49}$; Figure 36), but there was not a clear distinction between the functional and non-functional binding sites. After analyzing the change of the binding occupancy for these two factors, it appears that only some of these coincident binding events are instances of cooperative binding (Figure 37), where both Pho2 and Pho4 occupancy increases in response to phosphate starvation. 23 of 28 sites showing significant increase of both Pho2 and Pho4 are regulated by both factors ($p = 8.1 \times 10^{-16}$, Fisher's exact test). In contrast, when only one of the factors is significantly recruited in response to Pi limitation, the gene is not activated ('non-regulated' class in Figure 37, $p = 0.003$, Fisher's exact test). Thus it is these cooperative binding events that correlate with the gene activation by Pho4 ('regulated' class in Figure 37).

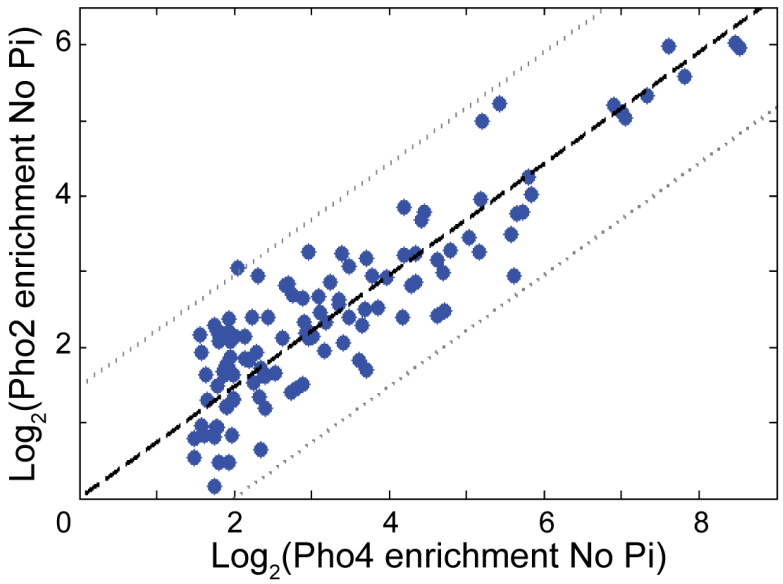


Figure 36. Correlation between Pho2 and Pho4 binding enrichment at Pho4 binding sites in no Pi conditions.

Scatter plot shows the binding enrichment of Pho2 and Pho4 over corresponding inputs for 115 Pho4-bound sites in no Pi conditions (blue dots). The black dashed line indicates the linear fit of the data points and the gray dashed lines indicate two units of $\log_2$(ratio) from the best fit. Pho2 and Pho4 binding occupancy are highly correlated with each other.

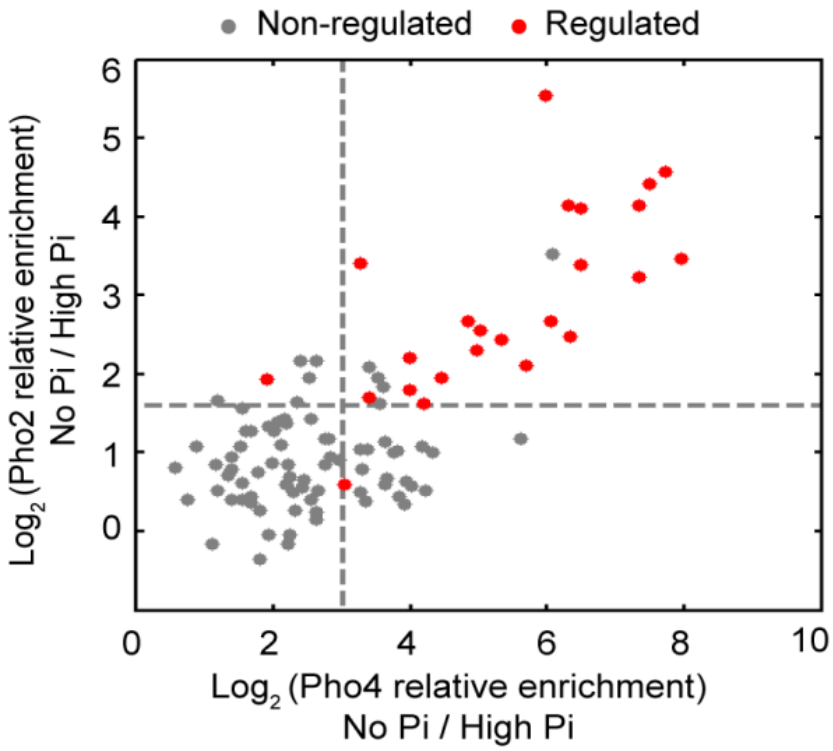


Figure 37. Recruitment of Pho2 and Pho4 after phosphate starvation.

Scatter plot shows the recruitment ($enrichment_{high}$ / $enrichment_{no}$) of Pho2 and Pho4 after Pi starvation for all Pho4-bound consensus sites within 800 bp of the transcription start site (TSS). 'Regulated binding sites' are sites associated with Pho4 regulated genes Figure 35. The increase in Pho2 and Pho4 binding after Pi starvation (recruitment) is calculated using the following equation: taking Pho4 as an example, $Enrichment_{No/High} = (Pho4_{ChIP\ No} / Pho4_{Input\ No}) / (Pho4_{ChIP\ High} / Pho4_{Input\ High})$. To estimate a threshold to define transcription factor recruitment, I calculated the mean and standard deviation of Pho2 and Pho4 recruitment at all consensus 'CACGTG' sites, excluding the sites at regulatory regions of Pho4 regulated genes (functional binding sites). I used mean + 2 standard deviations (s.d.) as the threshold to identify the sites showing the most significant recruitment and it is shown as the gray dashed lines.

Cooperative binding between two transcription factors often suggests that binding of one factor is dependent upon binding of the other factor. In high Pi conditions, Pho2 is constitutively localized within the nucleus (Huh et al., 2003), but binding of Pho2 is not observed at these cooperative sites (data not shown). The increase in Pho2 binding in no Pi conditions when Pho4 is nuclear localized indicates that Pho2 binding at these cooperative sites is dependent on Pho4 (Figure 37). To demonstrate that the binding of Pho4 is dependent on Pho2, I examined Pho4 binding occupancy in the *pho2Δ* strain in no Pi conditions. Recruitment of Pho4 to cooperative, regulated sites is entirely dependent on Pho2, whereas Pho4 binding to the non-cooperative, non-regulated sites is largely unaffected by the deletion of *PHO2* (Figure 38). Consistent with the change in recruitment, the occupancy of Pho4 (which is a more direct quantification of the amount of Pho4 molecules recruited at each binding site) at the non-cooperative sites was un-affected in the *pho2Δ* strain (Figure 39), in contrast to Pho4 occupancy at the cooperative sites. Further, these two classes of sites cannot be distinguished by their absolute occupancy, ruling out the possibility that the observed occupancy change at the cooperative sites in the *pho2Δ* strain is amplified by strong binding of Pho4 at these sites. Further, the reduction of Pho4 occupancy in the *pho2Δ* strains in no Pi conditions is correlated with the transcriptional activation contributed by the cooperative interaction between Pho2 and Pho4 (Co component) for Pho4-regulated genes ($R = 0.63$, $p = 0.0016$), consistent with a direct link between the cooperative binding of Pho2 and Pho4, and their cooperative component in gene regulation.

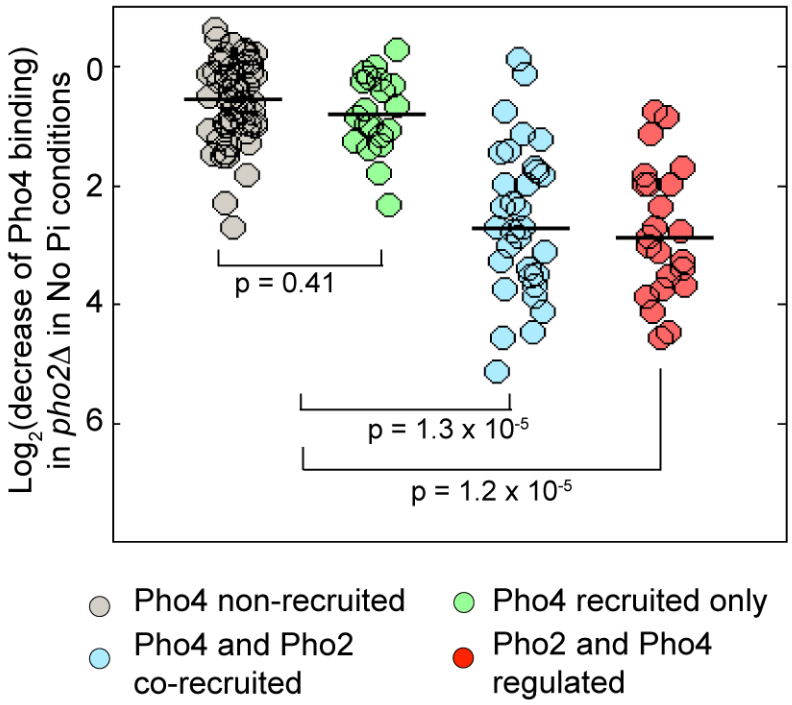


Figure 38. Change of Pho4 binding in *pho2Δ* strains at the Pho4-bound sites.

Plots show fold decrease in Pho4 occupancy at Pho4 bound high affinity binding sites in *pho2Δ* strains. Black lines indicate the median and P-values are calculated with two sample t-tests with unequal variance.

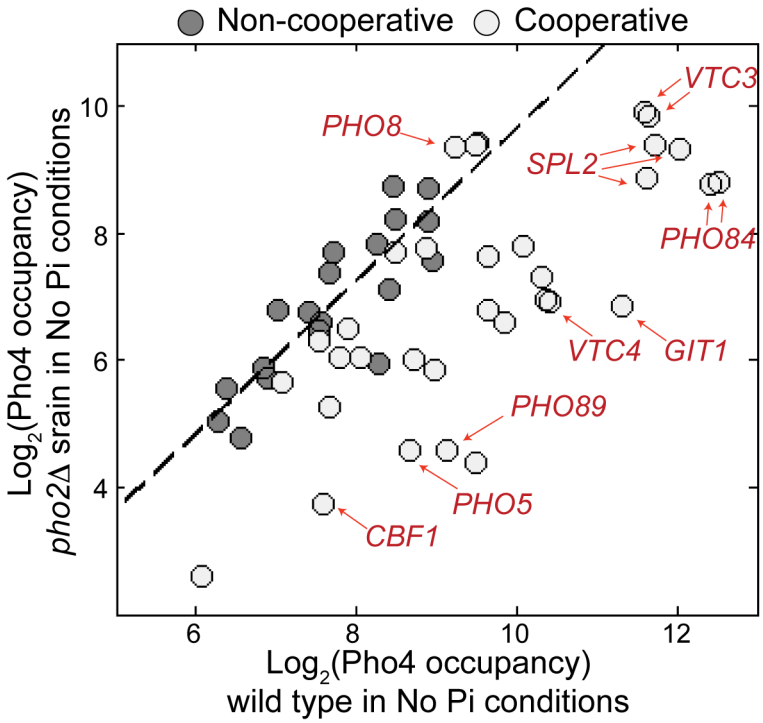


Figure 39. Correlation of pho4 binding occupancy between wild type and the *pho2Δ* strain in no Pi conditions

Plots show all consensus binding sites with significant Pho4 recruitment in the wild-type strain in no Pi conditions (Figure 36, points to the right of the gray vertical dashed line). ChIP-seq results Were normalized for the total number of alignable reads before comparison. Cooperative binding sites (light gray circles) and non-cooperative binding sites (dark gray circles) are defined as in Figure 36. The black dashed line indicates the best linear fit of non-cooperative binding sites with a $k = 1.2$. Pho4 occupancy decreases at cooperative binding sites in comparison to non-cooperative binding sites in *pho2Δ* strains in no Pi conditions. Binding sites of genes that can be activated by Pho4 alone do not show a decrease in Pho4 binding (*PHO8*), whereas binding sites of genes that are strictly dependent on both Pho2 and Pho4 show clear decreases in Pho4 binding (*PHO5*, *PHO89*, *GIT1*, *PHO84, VTC4, CBF1* etc).

Cooperative binding between transcription factors could happen in two ways, directly and indirectly. Two transcription factors could physically interact with each other when they bind to DNA; the DNA binding affinity of this protein complex as a whole is much larger than the affinity of either individual factor. The increase in DNA binding affinity stabilizes the transient association between these proteins and DNA, resulting in direct cooperativity. In another case, the binding sites of transcription factors may be restricted by nucleosomes and displacing the barrier nucleosome requires simultaneous binding of multiple transcription factors, presenting indirect cooperative binding (Adams and Workman, 1995; Miller and Widom, 2003; Vashee et al., 1998). To distinguish the mode of the cooperative binding between Pho2 and Pho4, I examined the local binding profiles of Pho2 and Pho4 binding at the cooperative and non-cooperative sites. For direct cooperative binding, I expected the ChIP signals of Pho2 and Pho4 to overlap at the binding sites; otherwise, the binding peaks should be mis-aligned at their respective binding locations. For the cooperative Pho4 binding events, the Pho2 and Pho4 ChIP signals are overlapping (Figure 40), while for the non-cooperative binding events, there is no juxtaposition of Pho2 and Pho4 ChIP signals (Figure 40). The close juxtaposition of Pho2 and Pho4 binding among the cooperative binding events suggests that this cooperativity is mediated through physical interaction between Pho2 and Pho4, as supported by biochemical evidence (Barbaric et al., 1996; Vogel et al., 1989).

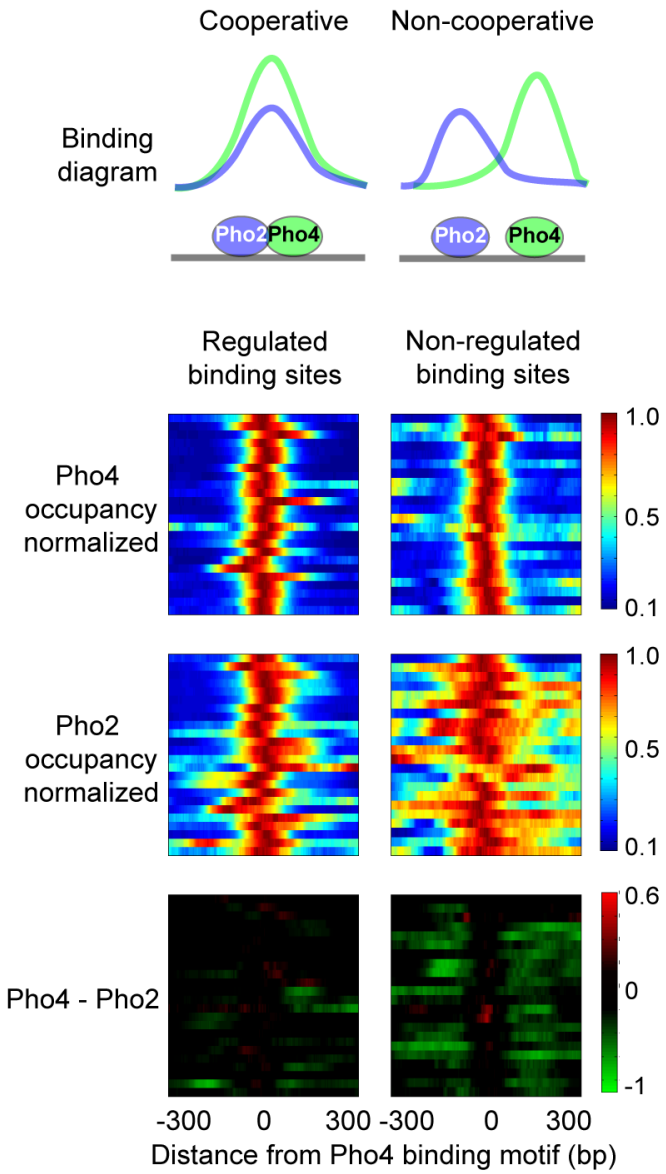


Figure 40. Local profiles of Pho2 and Pho4 binding at the regulated and non-regulated binding sites.

Heat map displaying normalized ChIP occupancy of Pho2 and Pho4 in no Pi conditions for all sites showing Pho4 recruitment (≥ mean + 2 s.d.).

Further, I analyzed the sequence motifs proximal to the cooperative and non-cooperative binding sites, as a means to identify the potential mechanisms that promotes this cooperative binding. Pho2 binds to 'AT' rich regions *in vivo* (Magbanua et al., 1997a) and prefers sequence motifs with alternating 'AA' 'TT' dinucleotide *in vitro (Zhu et al., 2009a)*. I used the PSSM of Pho2 (Table 5) to search for potential Pho2 binding sites near the high affinity binding sites of Pho4. The predicted Pho2 binding sites are enriched at a distance of 15 bp from the high affinity Pho4 binding sites for the cooperative and functional binding sites (Figure 41); on the other hand, the non-cooperative and non-functional binding sites showed depletion of Pho2 binding sequences at the same location (Figure 41), suggesting that the spatial organization of Pho2 and Pho4 binding motifs may promote cooperative binding of Pho2 and Pho4 and the ability to activate transcription.

Table 5. Position Specific Scoring Matrix of Pho2 binding *in vitro*.

The PSSM is determined in the study Zhu et al (Zhu et al., 2009b). The numbers in the matrix represent the frequency of observing the indicate nucleotide (raw) at the specified position (column) in a pool of 8-mer DNA sequences that are bound by Pho4. The most favored nucleotide at each position is highlighted in bold.

	Position in the DNA sequence							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
A	0.063551	0.198398	**0.767316**	**0.810313**	0.181954	0.198351	**0.780944**	**0.583919**
C	0.222198	0.015372	0.021460	0.004949	0.002784	0.012874	0.005286	0.130333
G	0.130333	0.005286	0.012874	0.002784	0.004949	0.021460	0.015372	0.222198
T	**0.583919**	**0.780944**	0.198351	0.181954	**0.810313**	**0.767316**	0.198398	0.063551

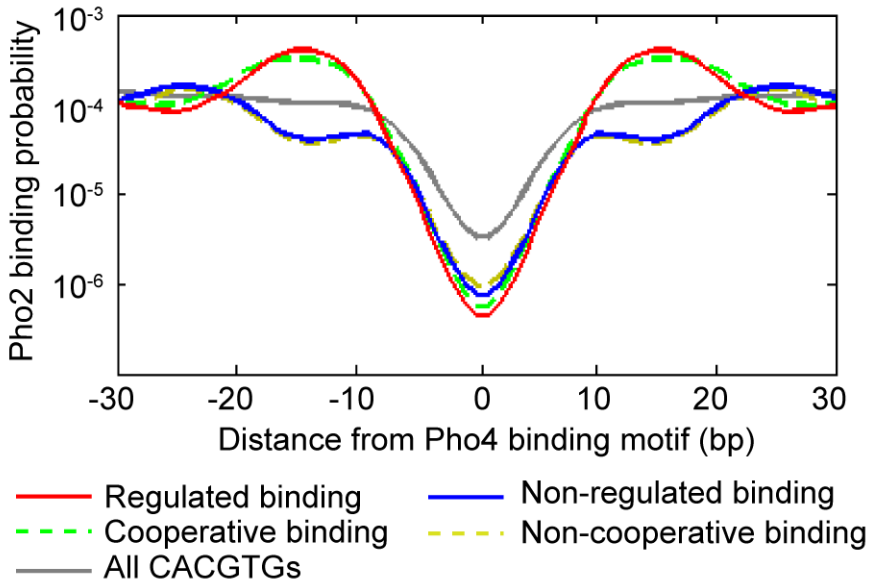


Figure 41. Enrichment of Pho2 binding motifs proximal to the Pho4 binding sites

Spatial organization of Pho2 binding motifs predicted based on *in vitro* binding specificity (Zhu et al., 2009a).

2.9. Prediction of Pho4 binding and function

Prediction of transcription factor binding and functional targets has been a challenging task, particularly for eukaryotic genomes. First, the binding specificity of only a small fraction of transcription factors is known (Bulyk, 2003). Second, most of the transcription factors recognize short sequence motifs (4-8 base pairs), which occur very frequently in large genome such as human and mouse where the regulatory elements could be located a few thousand base pairs away from the gene coding sequence. Last but not least, the *in vivo* binding of a transcription factor is influenced by a variety of biological components including chromatin structure, competition and interaction with other factors that are important for its binding affinity (Stormo and Fields, 1998).

To bypass some of these difficulties, computational methods have been developed to incorporate the information about evolutionary conservation in predicting the binding sites (Friberg, 2007; Zhao et al., 2012), under the assumption that the regulation of genes governing important biological processes should be conserved among closely related species and therefore the binding of transcription factors regulating the expression of these genes should also be conserved. However, these methods lack essential information regarding the mechanisms that determine where a transcription factor binds to in the genome and whether or not the binding event is functional.

Here I used experimental approaches to dissect the mechanisms that dictate the binding and regulation pattern of Pho4 at a whole genome scale, and aimed to test the idea of predicting transcription factor binding and regulation based an integrated mechanistic view. To develop a computational model, I took into account of all the determinants I found to influence Pho4

binding and regulation, and predicted all 'CACGTG' sites independent of evolutionary conservation, clustering of motifs, and relative positioning in the promoter.

2.9.1. Prediction of Pho4 binding

To predict Pho4 binding, I integrate information about Pho4 DNA binding preference, local nucleosome occupancy, and competition from Cbf1 into an equilibrium model (Granek and Clarke, 2005). At the equilibrium state, the probability of Pho4 binding to an 8-mer sequence at position *i* can be expressed as:

$$P_{bound,i} = \frac{[\text{Pho4}]}{[\text{Pho4}] + K_{d,\,i}}$$

Where $K_{d,\,i}$ is the equilibrium dissociation constant of the 8-mer sequence at position *i*, which can be derived from the position specific frequency matrix (PSSM) (Granek and Clarke, 2005). [Pho4] is the nuclear concentration of Pho4 and is set to have 0.5 binding probability at the optimal DNA binding sequence (Granek and Clarke, 2005; Maerkl and Quake, 2007).

From the equation above, the probability of the 8-mer sequence not being bound by Pho4 is

$$P_{unbound,\,i} = 1 - P_{bound,\,i} = \frac{K_{d,\,i}}{[\text{Pho4}] + K_{d,\,i}} = \frac{1}{K_{a,\,i} \times [\text{Pho4}] + 1}$$

Where $K_{a,\,i}$ is the equilibrium association constant at position *i* and is the inverse of $K_{d,\,i}$.

Restriction from nucleosomes and competition from Cbf1 can be reflected as variables that effectively decrease Pho4 binding probability or increase the probability of Pho4 being unbound.

I thus add two weighting terms representing these effects into the equilibrium binding probability model:

$$P_{unbound,\,i} = \frac{W_{Nuc,\,i} + W_{Cbf1,\,i} + 1}{W_{Nuc,\,i} + W_{Cbf1,\,i} + K_{a,\,i,\,adj} \times [\text{Pho4}] + 1}$$

$W_{Nuc,\,i}$ is given by equilibrium binding between DNA and nucleosomes lighted by a parameter k_{Nuc}:

$$W_{Nuc,\,i} = k_{Nuc} \times K_{a,Nuc,\,i} \times [\text{Nucleosome}]$$

Because

$$K_{a,\,Nuc,\,i} \times [\text{Nucleosome}] = \frac{[\text{DNA}]_{Nuc,\,bound,\,i}}{[\text{DNA}]_{Nuc,\,unbound,\,i}} = \frac{P_{Nuc,\,bound,\,i}}{1 - P_{Nuc,\,bound,\,i}}$$

$W_{Nuc,\,i}$ can be expressed as a function of the probability of a DNA motif being bound by nucleosomes ($P_{Nuc,\,bound,\,i}$). I assumed that this binding probability is proportional to the local nucleosome occupancy and it can be calculated as $P_{Nuc,\,Bound,\,i} = Occ_{Nuc,\,i} / Nuc_{Ref}$, where Nuc_{Ref} is an arbitrary number set as the maximum nucleosome occupancy in the genome. In the prediction, I use $Nuc_{Ref} = 1200$ since it gives a nucleosome binding probability close to 1 for the highest nucleosome occupied site in our dataset.

Similarly, $W_{Cbf1,\,i}$ is described as

$$W_{Cbf1,\,i} = k_{cbf1} \times \frac{P_{Cbf1\ Bound,\,i}}{1 - P_{Cbf1,\,Bound,\,i}}$$

The binding probability of Cbf1 can be described as $P = Occ_{Cbf1,\,i} / (Occ_{Cbf1,\,i} + Cbf1_{ref})$, where $Cbf1_{ref}$ describes the portion of free DNA not bound by Cbf1. I set $Cbf1_{ref} = 2000$ in this prediction; the model is largely insensitive to the $Cbf1_{ref}$ value (data not shown).

$K_{a,\,i,\,adj}$ is proportional to the equilibrium association constant and adjusted by a parameter k_{Pho4} to light the contribution of DNA binding preference of Pho4.

$$K_{a,\,i,\,adj} = k_{Pho4} \times K_{a,\,i}.$$

The detected transcription factor binding occupancy at a given genome location is likely contributed by all binding sites that are in close proximity. Thus, I include all potential Pho4 binding sites 40 bp up- and down- stream of the site of interest. The potential Pho4 binding sites are selected with the threshold that allows us to separate potential Pho4 regulatory elements from noise (Lam et al., 2008). If there are n potential Pho4 binding sites near position x on the genome, the binding probability at x is finally calculated as:

$$P_{bound,\,x} = 1 - \Pi\, P_{unbound,\,i}\ (i = 1, 2, 3 \ldots n)$$

I selected all 'CACGTG' sites on *S. cerevsiae* genome as the dataset to test the model prediction. I used our measurements of Cbf1 binding occupancy and nucleosome occupancy in no Pi conditions as inputs, and optimized the three parameters k_{Pho4}, k_{Cbf1} and k_{Nuc} to achieve the best prediction for Pho4 binding evaluated by area under the curve of a receiver operating characteristic (AUC-ROC). At $k_{Pho4} = 10$, $k_{Cbf1} = 2.5$ and $k_{Nuc} = 10$, 43 of top 50 and 60 of top 100 predicted binding sites are bound by Pho4, with AUC-ROC = 0.87. AUC-ROC changes less

than 0.15 and 0.12 for k_{Cbf1} and k_{Nuc} respectively, within two orders of magnitude of their optimal values. The model prediction is insensitive to k_{Pho4}.

In this model, nucleosome occupancy represents the occlusion effect of chromatin; Cbf1 occupancy represent the competitive binding influence of Cbf1 on Pho4; Pho4 DNA binding affinity takes into account Pho4 preference at flanking sequences of 'CACGTG' motif, which encodes the intrinsic preference about Pho4-Cbf1 competition. To explore how much each of these determining factors - nucleosome occlusion (Nuc), Pho4 preference at flanking sequences (Flank) and Cbf1 competition (Cbf1) - contribute to the prediction of Pho4 binding, I evaluated the predictive power of these factors separately as Ill as in combination (Figure 42 and Figure 43). For the model with each combination of determining factors, I first randomly sampled half of the dataset to fit the model for parameters that return the highest AUC-ROC value. Then I applied this best-fit parameter set to the model to predict Pho4 binding for all 'CACGTG' sites in the genome. I evaluated the model prediction with two indicators: AUC-ROC describing the overall predictive power (Figure 42), and accuracy as the number of correct predictions within the top 50 or 100 predicted targets (Figure 43). The variance and mean of the indicators can be estimated by repeating the prediction process multiple times.

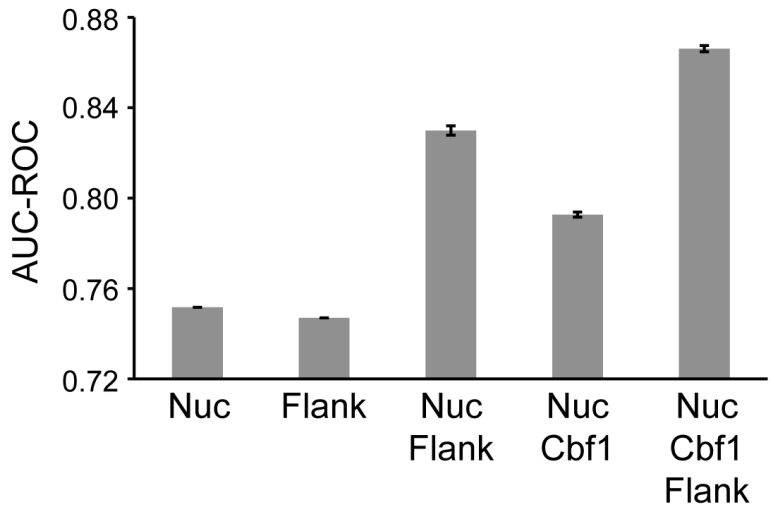


Figure 42. The Predictive power of the equilibrium model to predict Pho4 binding at all 'CACGTG' sites.

The equilibrium model predicts Pho4 binding at all 'CACGTG' sites with different combinations of *trans* effects. The area under the curve of receiver operating characteristic (AUC-ROC) represents the overall predictive power of models with different combinations of *trans* factors. AUC-ROC value is equal to 0.5 if a model does not predict Pho4 binding at all and is equal to 1 if a model has perfect prediction.

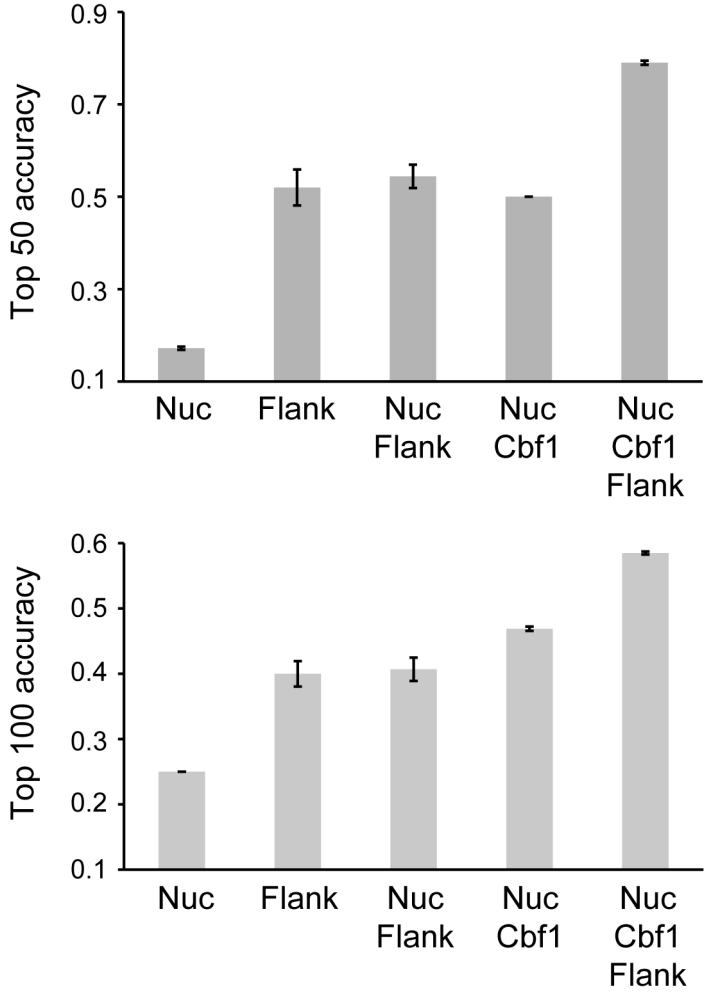


Figure 43. The accuracy of the equilibrium model to predict Pho4 binding at all 'CACGTG' sites.

Accuracy indicates the percentage of correct predictions within the top 50 or 100 predicted targets. Each column presents the mean of 10 prediction trials and the error bars indicate the standard deviation of the mean. The AUC-ROC value increases step-wise with addition of each of the three factors.

I observed that nucleosome occupancy and Pho4 preference at flanking sequences have similar overall predictive power for Pho4 binding (Figure 42), but they differ in the predictive accuracy of the top scored targets – prediction based on nucleosome occupancy alones has poor accuracy (Figure 43). Although Cbf1 binding itself does not predict Pho4 binding, its combination with nucleosome occupancy strongly improves the prediction accuracy (~200%), which is consistent with the model that Cbf1 determines Pho4 binding at nucleosome-free sites. Further, combining all three determining factors provides the best predictive power as well as the best accuracy among our tested models, and the AUC-ROC value shows a step-wise increase by adding each of the three factors. On the other hand, the accuracy of prediction is largely dependent on the preference of Pho4 at flanking sequence and the synergy between Cbf1 competition and local nucleosome occupancy (Figure 43, comparing Cbf1+Nuc with Cbf1 or Nuc, or comparing Flank+Cbf1+Nuc with Flank+Cbf1 or Flank+Nuc). Since both flanking sequence and Cbf1 binding contain information about Cbf1-Pho4 competition, this result suggests that Cbf1 competition has an important influence on Pho4 binding prediction.

In summary, incorporating the influence of *trans* effects into an equilibrium binding model, 43 of 50 (86%) (Figure 44) of the top predicted binding sites are indeed bound by Pho4 (AUC-ROC = 0.87). I find that all *trans* effects contribute significantly to the pattern of Pho4 binding in the genome. Either nucleosome occupancy or flanking sequences predicts Pho4 binding to high affinity consensus sites. However, if I consider only the group of top predicted targets, a prediction based on nucleosome occupancy or Cbf1 competition alone has poor accuracy, whereas the synergy between Cbf1 competition and nucleosome occlusion more accurately predicts binding to this group of target sites. I thus argue the necessity for understanding *trans*

influences, not only the restriction from nucleosomes, but also competitive binding factors, in achieving accurate computational prediction of transcriptional regulatory networks.

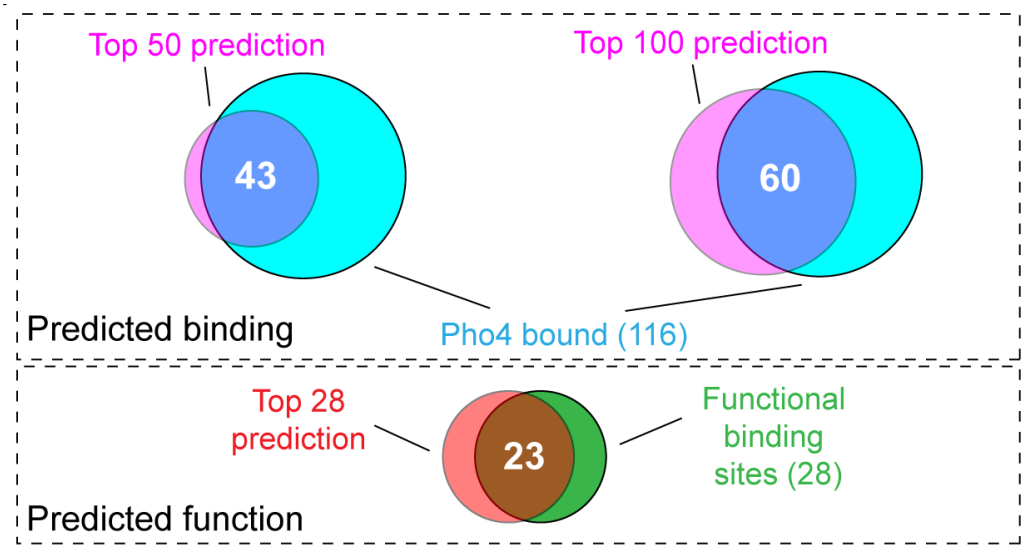


Figure 44. Prediction of pho4 binding and function

Venn diagram shows prediction of Pho4 binding and function derived from a model that incorporates competition and cooperativity.

2.9.2. Prediction of the functional outcome of Pho4 binding events

Of the 115 experimentally-determined Pho4 binding sites, fewer than 25% are able to promote activation of gene transcription in no Pi conditions. Another *trans* effect - the cooperative interaction with Pho2 – determines whether Pho4 binding is functional, leading to transcriptional activation. To predict the functional binding events for Pho4, I take into account the cooperative recruitment of Pho2 and Pho4 by calculating the geometric mean (square root of the product) of Pho2 and Pho4 recruitment after Pi starvation. 23 of the top 28 ranked binding sites predicted to be functional are actually associated with transcriptional activation by Pho4 (AUC-ROC = 0.992). In detail, 20 of 23 promoter binding events are associated with genes regulated by Pho4 from mutant cycle analysis; 3 of 5 non-promoter binding events are linked to Pi dependent anti-sense transcripts (*B. Zid et al., unpublished*). One of these anti-sense transcripts regulates the expression of Kcs1, a key enzyme in the phosphate starvation signaling pathway (Nishizawa et al., 2008). The other two, at the location of the genes *SFK1* and *GTO1*, have not been previously identified and are associated with Pho4 binding at the 3' end of the gene coding sequence (Figure 45). *SFK1* appears to be of particular interests – the anti-sense transcript is dramatically induced by phosphate starvation, the sense transcripts are inhibited by phosphate starvation by ~ 3 fold in a Pho4-dependent fashion (and likely due to the induction of the anti-sense transcript), and this anti-sense transcript overlaps partially with the 5' UTR of *SFK1* (Figure 45, upper panel), suggesting possible inhibition of the translation of Sfk1. Sfk1 is a plasma membrane protein and influences the localization of Stt4p, a phosphatidylinositol-4-knase that functions in the Pkc1 protein kinase pathway (Audhya and Emr, 2002; Audhya et al., 2000; Yoshida et al., 1994). The observed induction of anti-sense and down-regulation of *SFK1*

transcripts may indicate potential crosstalk between PHO pathway and PKC signaling pathway during phosphate starvation.

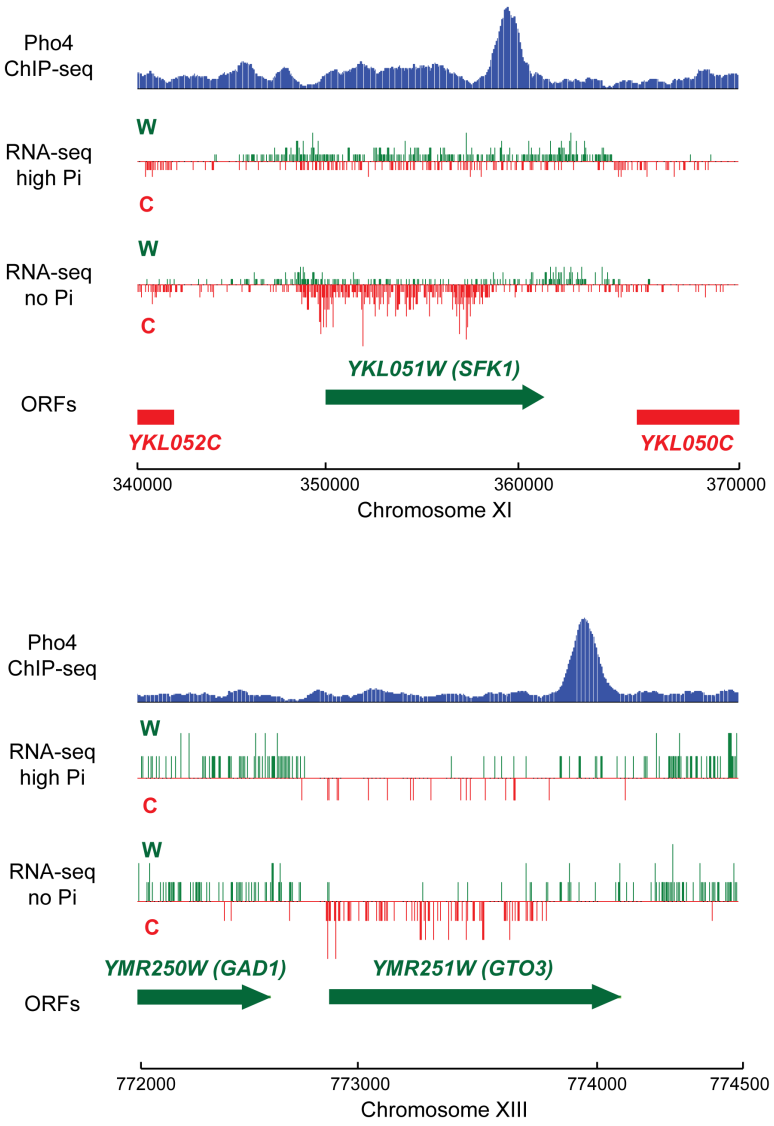


Figure 45. Two predicted functional binding sites associated with anti-sense transcripts.

The plots show Pho4 binding occupancy in no Pi conditions (blue), Counts of RNA-seq reads on the forward (W, green) and reverse (C, red) strains in high Pi and no Pi conditions. The coding regions of open reading frames are labeled with green and red arrows to represent the direction of the transcripts. Anti-sense transcripts of *YKL051W* and *YMR251W* show significance increase in response to phosphate starvation (compare the transcripts on the Crick strands for both genes). RNA-seq data is kindly provided by *Brian Zid*.

To explore the possibility of predicting transcriptional outcome from DNA sequence information, I examined how the spatial relation between Pho2 and Pho4 binding sites predicts the functional outcome of Pho4 binding events. I calculated the Pho2 binding probability in 10 bp windows, centered 5, 10, 15, 20, 25, 30, and 35 bp away from the Pho4 binding site (Figure 46) and used an AUC-ROC value to evaluate the predictive power. The prediction peaks at 15 bp from the Pho4 binding site (AUC-ROC = 0.725) and then sharply drops to the level of non-predictive when the distance between the Pho4 binding site and the center of the window deviates from 15 bp. This result suggests that the presence of Pho2 binding sites 15 bp away from Pho4 binding site promotes transcriptional activation and is highly predictive of the functionality of the Pho4 binding events. Together with our data showing the cooperative binding between Pho2 and Pho4 at functional binding sites, and examples of physical interaction between these two factors at *PHO5* promoter (Vogel et al., 1989), I conclude that the spatial arrangement of Pho2 and Pho4 binding sites is critical and predictive for the functional outcome of Pho4 binding events.

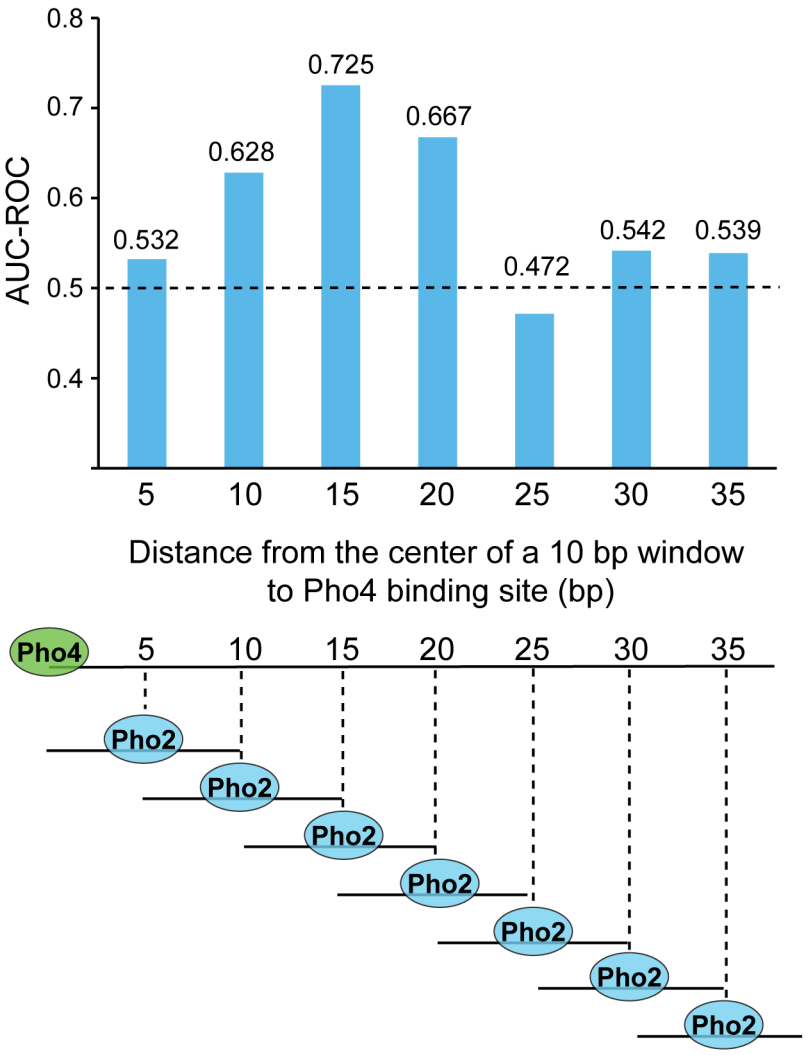


Figure 46. Sequence-based prediction of the functional outcome of Pho4 binding events.

The chart presents the AUC-ROC value of the model prediction based on the Pho2 binding probability calculated from a series of 10-bp window DNA sequences centered every 5 bp from the Pho4 binding sites. The diagram below illustrates the sequence windows corresponding to each of the columns in the chart.

3. Discussion

3.1. Specific transcriptional control of the PHO pathway in *S. cerevisiae*

How does Pho4 specifically regulate only ~30 genes in the genome? In high Pi conditions, Pho4 is phosphorylated and exported to the cytoplasm (O'Neill et al., 1996; Schneider et al., 1994). Competition from Cbf1 increases the threshold of transcriptional activation by Pho4, preventing spurious activation by low levels of nuclear Pho4. In no Pi conditions, phosphorylation of Pho4 is inhibited (O'Neill et al., 1996; Schneider et al., 1994) and the increase in the concentration of unphosphorylated nuclear Pho4 enables it to compete effectively with Cbf1 at sites where Cbf1 is likely bound – those with certain flanking sequences found in genes within the phosphate regulon; inappropriate binding and activation of 'CACGTG' containing genes outside the phosphate regulon is prevented. The dynamic range of binding and gene activation is increased as a result of the reduction in Pho4 binding due to Cbf1 competition in high Pi conditions and its condition-dependent cooperative interaction with Pho2 in no Pi conditions (Pho2 interacts only with unphosphorylated Pho4 during phosphate limitation (Komeili and O'Shea, 1999)).

Stress and changes in nutrient availability trigger high levels of induction of specific genes that were tightly repressed in the absence of the perturbation. Moreover, this induction is commonly carried out by transcription factors belonging to families in which the members have similar DNA binding specificity (Robinson and Lopes, 2000). What mechanisms are used to ensure that only the correct genes are activated to high levels, and how is expression of these genes kept low in the absence of stress? The phosphate regulon uses competition to prevent spurious activation by low levels of nuclear Pho4 when PHO pathway induction is not necessary;

and employ both competition and cooperativity to ensure that only phosphate-responsive genes are activated in response to phosphate limitation. The dynamic range of binding and gene activation is increased as a result of the reduction in Pho4 binding due to Cbf1 competition in high Pi conditions and its condition-dependent cooperative interaction with Pho2 in no Pi conditions (Pho2 interacts only with unphosphorylated Pho4 during phosphate limitation (Komeili and O'Shea, 1999)). This activation is not "leaky" and is of high dynamic range. Competition comes from nucleosomes, which restrict access of Pho4 to many 'CACGTG' sites not associated with genes in the phosphate regulon. Competition also comes from Cbf1, a transcription factor that has DNA binding specificity similar to that of Pho4. Other stress- and nutrient response transcriptional programs may employ similar strategies to trigger high levels of induction of specific genes that are tightly repressed in the absence of the perturbation.

3.2. Roadmap for understanding the specific transcriptional control in the eukaryotic genome

The transcriptional control in other eukaryotic genomes tends to be more complicated. The *Saccharomyces cerevisiae* genome contains ~6000 genes, of which 3% - 4% encode proteins functioning as DNA binding factors (Harbison et al., 2004). In the human and mouse genomes, about 10% of the genes are expected to encode DNA binding proteins, many of which are predicted to be transcription factors (Lander et al., 2001; Venter et al., 2001). The large number of transcription factors hints at multiple-layer and complex transcriptional regulatory networks. This complexity imposes more challenges for the specific regulation by transcription factors. For instance, the number of transcription factors increases more than 10 times from yeast to human, but the number of transcription factor structure families remains almost the same, suggesting that more transcription factors recognize the same DNA sequences in these large genomes. In addition, regulatory elements in human genome can be located distally from the gene coding region, which decreases the specificity provided by a given binding motif. It is not known how transcription factors in human or mouse specifically regulate the transcription of a defined set of genes. The key to this problem may involve mechanisms ranging from chromatin accessibility, temporal and cell type specific expression of partner and competitive factors, to the molecular details of the binding preference and interactions of factors.

The work presented here provides a general road map to dissect the components that determine specific transcriptional control in a regulatory system. In the PHO pathway of *S. cerevisiae*, three layers of regulation are critical for the specific regulation of PHO genes in response to phosphate starvation: the first layer of regulation comes from the restriction of nucleosomes, where the nucleosome-occluded sites are not accessible for Pho4 binding; the

second layer of regulation comes from the competitive binding of Cbf1, which determines Pho4 binding at the nucleosome-free sites; the third layer comes from cooperativity with Pho2, where the cooperative binding between these two factors specifically happens at the genes regulated by Pho4.

These layers of regulation may emerge as important determinants for the regulation specificity of transcription factors in higher eukaryotes, in spite of increased complexity. For example, the restriction of nucleosomes may not only originate from directly competition between nucleosomes and transcription factors, but also come from the barrier of the heterochromatin. The influence of heterochromatin may be extensive in vertebrates, whereas it is limited to the telomeric regions and the mating locus in budding yeast (Miele et al., 2009). In addition, the restriction of nucleosome may play different roles in transcription factor binding - rather than to block the binding site, it is possible that nucleosome imposes a higher activation threshold so that the binding of transcription factors requires a cluster of the binding sites. Understanding how the binding and regulation of transcription factors is influenced by heterochromatin structure as well as the direct competition from histones will shed light on the molecular mechanisms of specific transcriptional activation. In the second layer, it is possible that the competition exists in mammalian systems due to the high similarity among the binding sequences of transcription factors (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). The competition between multiple factors with similar binding specificity provides a plausible explanation for dosage responses that are commonly observed for transcriptional regulators, whose reduced expression causes defective induction in its cognate pathway and over-expression introduces crosstalk with other responses. For example, during the early stages of T-cell commitment, a key regulatory GATA-3 is necessary to differentiate T-cell precursors from stem

cells (Hattori et al., 1996; Ho and Pai, 2007), but forced expression GATA-3 turns on the transcription programs of other cell types (Hosoya et al., 2009). Examining the possible competition between transcription factors may shed light on the mechanisms underlying these observations and provide insights into the determinants of cell type specification in this and other similar systems. In the third layer, the co-recruitment of multiple transcription factors may be essential for transcriptional regulation. For instance, simultaneously recruitment of several key transcriptional regulators is essential for the regulation of genes critical for maintaining the pluripotency of stem cells (Goke et al., 2011; Young, 2011). Dissecting the roles of each of these co-recruited factors and uncovering the underlying mechanisms will be critical to advance our knowledge and connect the physical binding behavior with its physiological regulation outcome.

In mammalian cells, transcription factors within the same structural family commonly recognize similar DNA motifs but regulate diverse biological processes (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). It is unclear how faithful and distinct regulation is achieved with transcription factors of overlapping specificity. Our results suggest that chromatin structure influences the accessibility of potential binding sites and competition among transcription factors can facilitate discrimination of subtle differences in DNA binding specificity across the genome. Cooperation with conditionally, spatially or temporally expressed co-activators (Sharrocks, 2001) may provide additional selection for binding events to elicit specific transcriptional outcomes. Moreover, competition among transcription factors may have significant effects on transcriptional regulation, controlling the threshold for gene activation and limiting potential crosstalk between different signaling pathways. Overall, our findings suggest that the specificity of transcriptional regulation is a composite of the DNA binding specificity,

nuclear abundance and functional interactions of each transcription factor in the context of eukaryotic genome architecture. Understanding the grammar of specific regulation of individual transcription factors will provide insights into decoding the complex regulatory network of the eukaryotic genomes.

3.3. Pho4 binding at non-consensus (low affinity) binding sites

Most of my study has focused on the binding sequences with highest affinities with Pho4. However, other 8-mer sequences have also been shown to regulate Pho4-dependent activation of the PHO genes, such as the low affinity binding sites at the promoters of *PHO5* and *PHO89* . To examine if the low affinity sequences also contribute significantly to Pho4 binding, I expanded my analysis to additional 1758 low affinity binding sites, whose Pho4 binding affinity pass 0.0003 of the PSSM score, a threshold determined by the verified functional low affinity binding sites at the promoters of *PHO5* and *PHO89* (Barbaric et al., 1992; Munsterkotter et al., 2000; Ogawa et al., 1995; Venter et al., 1994). Only 34 of them (2%) are bound by Pho4 using the same threshold I applied to high affinity binding sites. 9 of the top 11 binding events are likely due to Pho4 binding to an adjacent high affinity binding site (within 200 bp), including the low affinity binding sites at promoters of *PHO5*, *PHO89*, *PHO84*, *GIT1*, etc. Excluding the sites where high affinity sites are positioned nearby, Pho4 binds to only ~1% (20 sites) of the low affinity sites, which is only slightly higher than the occurrence estimated by chance (0.6%). 7 binding events at low affinity motifs are associated with Pho4 dependent transcriptional activation; 6 of them are near a consensus high affinity binding site. These results together with studies on individual promoter (Kim and O'Shea, 2008; Lam et al., 2008) suggest that low affinity sites mainly function in conjunction with high affinity sites to regulate gene expression.

Pho4 binding at low affinity sites is also highly dependent on nucleosome occupancy: none of the 521 nucleosome occluded low affinity sites are bound by Pho4 and 76% (26 of 34) of the Pho4 bound low affinity sites are accessible. Competition between Pho4 and Cbf1 can be extended to low affinity sites as well, as none of the 102 ‘TCACGTGA’ motifs (the only low affinity sequence containing ‘CACGTG’ core motif) in the genome are bound by Pho4 while 91

are bound by Cbf1. After C*BF1* deletion, Pho4 binding at 'TCACGTGA' sites increases dramatically.

Overall, only rare binding events (20 of 1758) are observed at sequences deviating from the consensus binding site 'CACGTG' by more than one nucleotide; these binding events have similar dependence on chromatin structure and Cbf1 competition.

3.4. Comparison of Pho4 and Cbf1 binding

The competitive binding between a transcription factor and nucleosomes is combinatorially influenced by the concentration of transcription factor in the nucleus, the binding affinity of DNA sequences with transcription factor, and the stability of the nculeosomes covering the binding site.

Cbf1 binding is highly enriched at accessible Pho4 binding sites (207 of 248 accessible sites are bound by Cbf1) and is less enriched at inaccessible Pho4 binding sites (83 of 216 inaccessible sites are bound by Cbf1). The instances of sites bound by both Cbf1 and nucleosomes can be interpreted as binding sites occupied by Cbf1 in a fraction of the population and occupied by nucleosomes in another fraction, as ChIP-seq and nucleosome mapping report on population-Iighted averages.

Cbf1 and Pho4 apparently differ in their ability to bind nucleosome-occupied sites - Cbf1 can bind to some of these sites, but Pho4 cannot. In addition, of the nucleosome occupied sites that are also bound by Cbf1, I did not generally observe Pho4 binding to these sites even in the absence of Cbf1. The binding affinity of Pho4 and Cbf1 to their most preferred sequence is similar, 11.1 nM and 16.6 nM respectively (Maerkl and Quake, 2007). The difference in the ability of Cbf1 and Pho4 to bind nucleosome occluded sites may be explained by their difference in nuclear concentration: Cbf1 is at much higher concentration (6890 copies/cell based on the quantification of western blot (Ghaemmaghami et al., 2003), or 5000 copies/cell assuming 5×10^7 molecules total protein per cell, estimated from proteomic mass spectrum (de Godoy et al., 2008)) inside the nucleus than is Pho4 (~100 copies/cell (de Godoy et al., 2008)).

3.5. Differential influence of Cbf1 competition based on promoter architecture

Most of the PHO genes can be classified into two categories based on the relative position of Pho4 binding sites and promoter nucleosomes. In the first group, at least one high affinity Pho4 binding site is located in an accessible region (nucleosome-free or the linker region between adjacent nucleosomes); *PHO84*, *SPL2*, *PHO8*, *VTC1* and many PHO genes belong to this category. In the second group, the gene promoter contains a nucleosome-occluded high affinity binding site and an adjacent nucleosome-free low affinity binding site; only 5 PHO genes in budding yeast genome belong to this category, *PHO5*, *PHO11*, *PHO12*, *PHO89* and *PHM6*. These different types of promoter architecture exert significant influence on the activation threshold of a gene and the dynamics of inducible transcription (Kim and O'Shea, 2008; Lam et al., 2008).

Similarly, the competition between Pho4 and Cbf1 may regulate the activation of PHO genes differentially based on their respective promoter architecture – the competition from Cbf1 would exert stronger influence at the accessible binding sites than at the inaccessible binding sites, as the influence of competition is likely to be reduced by the "protection" from nucleosomes. Intriguingly, the fold induction of *PHO5* (4.2-fold) and *PHO89* (4.8-fold) in the *cbf1Δ* strain in high Pi conditions is larger than that of *PHO8* (1.4-fold) and *VTC1* (3.1-fold) (Figure 30). However, if the induction in the *cbf1Δ* in high Pi conditions is viewed as a percentage of the induction in wild type in no Pi conditions, *PHO8* and *VTC1* are induced to a much higher level (45% and 35%, respectively) than are *PHO5* and *PHO89* (4.5% and 11%, respectively), indicating that the promoters with accessible Pho4 binding sites are more susceptible to the influence of transcription factor competition.

3.6. Specification of phosphate and methionine starvation responses

Both Pho4 and Cbf1 bind to 'CACGTG' core sequences with substantial overlapping at the Pho4 binding sites. Yet they respond to different environmental signals and activate distinct transcription programs – phosphate starvation response and the methionine starvation response. How would the transcriptional regulatory network of each pathway ensure the nutrient-specific response when the binding of their key transcription factors does not provide sufficient specification?

In the PHO pathway, Pho4 directly activates the transcription of PHO genes in the presence of Pho2 during phosphate starvation; both Pho2 and Pho4 contain a DNA binding domain as well as an acidic activation domain that is responsible for recruiting chromatin remodelers and components of the transcription initiation complex (Magbanua et al., 1997b; McAndrew et al., 1998; Svaren et al., 1994). The pathway activity is mainly controlled by the localization of Pho4 and its interaction with nuclear localized Pho2 (Komeili and O'Shea, 1999). In the MET pathway, Cbf1 induces the expression of MET genes together with the MET transcription factors, Met4, Met28, Met31 and Met32; the DNA binding capability and activation function are decoupled among these factors (Blaiseau and Thomas, 1998; Kuras et al., 1996; Lee et al., 2010; Thomas et al., 1992). Met4 is the sole activator of the sulfur metabolic pathway but requires interaction with the other MET transcription factors to associate with DNA. The pathway activity is regulated by the protein stability of the activator, Met4 (Rouillon et al., 2000). The difference in the structure of these two regulatory networks may contribute to their specific regulation of transcription programs. First, lack of activation capability in Cbf1 allows its inhibitory influence on the PHO genes and avoids broad activation of the genes to which it binds. Second, the combinatorial regulation by Met28, Met31, Met32, and Cbf1 allows specific

recruitment of Met4 at the promoters of MET genes, keeping PHO genes from being activated merely due to Cbf1 binding. The combinatorial regulatory structure of MET pathway may provide the plasticity necessary to separately evolve the phosphate and sulfur specific metabolic responses. Third, the strong competition from Cbf1 over the 'T-CACGTG' sites protects the MET genes from the activation by Pho4 under phosphate starvation conditions, preventing potential crosstalk between the two pathways. The exceptional conservation of 'T-CACGTG' sequences at the promoters of MET genes across multiple fugal species argues strongly for this possibility and an evolutionary force to separate the phosphate and sulfur starvation responses. Last but not least, the cooperativity between Pho2 and Pho4 further refines the specific PHO regulon, where lack of this cooperativity causes much broader activation in the genome (data not shown). Overall, the pathway specific regulation is likely to be collaborative efforts from specification at different levels of the regulatory networks, ranging from the complex interactions among multiple transcription factors, to the modular function of individual proteins, to the molecular details of biophysical properties. Further studies to examine the specification and crosstalk between multiple pathways that share similar regulatory sequences will shed light on the mechanisms of specific transcription control in the context of all the regulators in eukaryotic genome. Further, exploring the possible relationship between having specific transcriptional control and fitness advantage will provide us with exceptional opportunities to study the interaction between metabolic pathways in the course of evolution.

4. Experimental Procedures

4.1. Strains

All yeast strains used in this study except EY2628 Were constructed from EY57 (K699 *MAT***a** *ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 ura3*) (Table 6). C-terminal Avi-tagged strains were constructed through PCR-based integration (Bahler et al., 1998). DNA sequence encoding the C-Avitag (GLNDIFEAQKIEWHW) was inserted into the pFA6a-GFP-TRP1 plasmid between the *Pac*I and *Asc*I restriction sites to replace the GFP coding sequence. Standard PCR primers (Bahler et al., 1998) were used to amplify the *C-Avitag-TRP1* cassette and integrate it at the C-terminus of the targeted genes at their native loci. N-terminal Avi-tagged strains Were constructed through the pop-in/pop-out gene replacement method (Guthrie and Fink, 1991). Briefly, a *CgURA3* cassette was first inserted between the native promoter and the second codon of the targeted genes (pop-in). DNA sequence encoding the N-terminal Avitag (MSGLNDIFEAQKIEWHEGAPGSGS) was then integrated to replace the inserted *CgURA3* cassette by selecting for uracil auxotrophy (pop-out) (Ausubel, 1987). To stably integrate *E. coli* biotin ligase into the *S. cerevisiae* genome, the fragment containing *E. coli birA* was digested from pRS313-BirA-NLS plasmid (Diagenode) (van Werven and Timmers, 2006) with *Not*I and *Spe*I, and then ligated into the pRS306 plasmid at the same restriction sites. pRS306-BirA-NLS plasmid was linearized at the *Stu*I restriction site and integrated at the native *URA3* locus of all Avi-tagged strains. Pho2 and Pho4 Avi-tagged proteins function properly in activating *PHO* gene transcription in response to Pi starvation (data not shown). The Cbf1 Avi-tagged strains show normal growth rate and methionine prototrophy. All gene deletion strains Were constructed with pFA6a or pUC based plasmids with various selection markers by standard PCR integration methods.

4.2. Media and growth conditions

Phosphate-free synthetic complete medium was prepared from Difco phosphate-free Yeast Nitrogen Base and supplemented to final concentration of 2% glucose, 1.5 mg/ml potassium chloride, 0.1 mg/ml sodium chloride and amino acids, as described previously (Lam et al., 2008). Monobasic potassium phosphate was added to phosphate-free medium to make high phosphate (Pi) medium containing a final concentration of 10 mM inorganic phosphate. All media Were adjusted to pH 4.0 with HCl (Thomas and O'Shea, 2005). Yeast strains Were grown at 30 °C with shaking and cell samples Were collected at early/mid-logarithmic phase (OD_{600} 0.3 ~ 0.4).

To induce the Pi starvation response, yeast cells were first grown in 10 mM Pi medium to early/mid-logarithmic phase. Cells Were then harvested by filtering and washed 2 – 3 times with no Pi medium pre-warmed to 30 °C. Finally, cells were resuspended in pre-warmed no Pi medium and grown at 30 °C for 1 hour.

Table 6. List of yeast strains used in this study.

Strains	Genotype
EY 57	WT
EY 134	*Pho80::HIS3*
EY 219	*Pho4::TRP1 pho80::HIS3*
EY 337	*pho2::CgLEU2*
EY 338	*pho2::CgLEU2 pho4::CgTRP1*
EY 692	*pho4::PHO4-GFP ADE2*
EY 1519	*cbf1::CgLEU2*
EY 1710	*pho4::CgURA3*
EY 1801	*cbf1::CgLEU2 pho80::HIS3*
EY 2592	*pho4::PHO4-C-Avitag-TRP1 ura3::pRS306-birA*
EY 2593	*pho2::N-Avitag-PHO2 ura3::pRS306-birA*
EY 2596	*ura3::pRS306-birA*
EY 2597	*cbf1::CBF1-C-Avitag-TRP1 ura3::pRS306-birA*
EY 2599	*pho4::PHO4-C-Avitag- TRP1 pho2:: HI3MX6 ura3::pRS306-birA*
EY 2628	*pho4::PHO4-GFP cbf1::HIS3MX6 ADE2*
EY 2630	*pho4::CgURA3 cbf1::CgLEU2*
EY 2633	*tye7::HIS3MX6 rtg3::TRP1*
EY 2680	*Pho4::TRP1 pho80::HIS3 cbf1::CgLEU2*
EY 2681	*pho4::PHO4-C-Avitag-TRP1 pho80::HIS3MX6 ura3::pRS306-birA*
EY 2682	*pho4::PHO4-C-Avitag-TRP1 cbf1::HIS3MX6 pho80::CgLEU2 ura3::pRS306-birA*

4.3. Defining the consensus Pho4 binding motif

High affinity Pho4 binding motifs were determined using the position specific scoring matrix (PSSM) described in (Lam et al., 2008), which was derived from *in vitro* measurement of Pho4 DNA binding affinities (Maerkl and Quake, 2007). I selected the most stringent threshold (0.0075) to recapitulate *in vivo* validated high affinity binding sites as the threshold of Pho4 high affinity binding motifs (Lam et al., 2008). Since all determined high affinity binding motifs at this threshold contain 'CACGTG' as core sequence, I define these motifs as the consensus 'CACGTG' binding motif for Pho4. All 'NCACGTGN' motifs except 'TCACGTGA' in *S. cerevisiae* genome meet the threshold.

4.4 Biotin-tagging immuno-precipitation with high throughput sequencing (Bio-ChIP-Seq)

Bio-ChIP-Seq was modified from techniques previously described (Kolodziej et al., 2009; Lam et al., 2008; van Werven and Timmers, 2006). ~100 OD units of cells were collected for high Pi and no Pi conditions. Cells Were crosslinked with 1% formaldehyde for 15 minutes and then quenched with 125 mM glycine for 5 minutes at room temperature. Cells were collected by centrifugation, immediately washed with cold PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Sodium phosphate dibasic, 2 mM potassium phosphate monobasic, pH 7.4), and mechanically lysed with glass beads in lysis buffer (50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate). Chromatin was fragmented by sonication of the lysate and the supernatant was pre-cleared with protein G dynabeads (Invitrogen) for 2 hours at 4 °C. M-280 dynabeads (Invitrogen) Were blocked with lysis buffer containing 1% cold fish skin gelatin (Sigma-Aldrich) at 4 °C for 2 hours and then incubated with pre-cleared cell lysate overnight at 4°C. After incubation, the dynabeads Were washed with lysis buffer, high salt wash

buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate), lithium wash buffer (10 mM Tris/HCl, pH 8.0, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.1% Na-Deoxycholate), and SDS wash buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA, 3% SDS) for 2 × 2 minutes at room temperature, and 1× 2 minutes with TE buffer (10 mM Tris/HCl, pH 8.0, 1 mM EDTA). Crosslinks Were reversed by incubation of samples at 65 °C for at least 6 hours in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.8% SDS. RNA and proteins in the samples Were digested with 20 ug/ml RNase at 37 °C for 2 hours and 0.2 mg/ml proteinase K for 2 hours at 55 °C. DNA was then purified with phenol-chloroform extraction and precipitated with ethanol. 5% of the volume of cell lysate was removed after sonication and used to prepare the input DNA for each ChIP experiment. ChIP DNA concentration was estimated with the Pico-green DNA detection kit (Invitrogen). Sequencing libraries were prepared following the Illumina protocol and libraries with size between 200 to 300 bp were selected for PCR amplification and assayed with an Agilent bioanalyzer. Libraries were sequenced with an Illumina Genome analyzer II, and 36 base sequence tags were aligned to the *S. cerevisiae* genome with ELAND. On average, 2.9 and 12.5 million uniquely aligned DNA sequencing reads Were obtained for ChIP and input samples, respectively. The uniquely aligned reads Were mapped on the genome and extended 80 bp from the read start position to cover the average length of inserted DNA between sequencing adaptors as determined by the Agilent bioanalyzer.

In summary, I performed Bio-ChIP-seq experiments on Pho2, Pho4, Cbf1 and mock samples in both high and no Pi conditions (60 minutes after Pi starvation), and on Pho4 in a *pho2Δ* strain grown in no Pi conditions, and in *pho80Δ* and *cbf1Δ pho80Δ* strains grown in high Pi conditions. Sequencing libraries were prepared for both ChIP DNA and input DNA (from the supernatant of total cell lysate) following the Illumina protocol. Libraries with size between 200 bp to 300 bp

were selected for PCR amplification and sequenced with an Illumina Genome analyzer II. 36 base sequence tags were aligned to the *S. cerevisiae* genome with ELAND. On average, 2.9 and 12.5 million uniquely aligned reads were obtained for ChIP and input samples, respectively.

4.5. Identification of ChIP binding regions

I normalized the input sequencing data against the background of the corresponding ChIP sequencing data, excluding the binding regions that are over-represented in the ChIP data. I first used ChIP sequencing data to identify putative binding regions with a false discovery rate of 0.05 using a simulated randomly distributed genome background. Second, I calculated a scaling factor for each chromosome by linear regression of the ChIP data against the corresponding input data in 1000 bp windows, excluding the identified putative binding regions. The input results were then scaled by the scaling factors for each chromosome and compared with the ChIP data to quantify the enrichment of putative binding regions. For each putative region, I also calculated the statistical significance (P-value) of Pho4 enrichment relative to input based on binomial distribution and corrected for multiple hypothesis testing with Benjamini-Hochberg correction (Q-value).

4.6. Determination of binding at consensus binding sites

Briefly, I calculated a 20 bp window average centered on every CACGTG motif for ChIP results ($mean_{ChIP}$) and Mock IP results ($mean_{mock}$), and a 150 bp window average for the normalized input results ($mean_{Input}$) to minimize local background variations. I score a signal as binding at a given consensus CACGTG site if it meets all of the following requirements: (1) significantly bound (ChIP occupancy, $p \leq 0.05$), (2) significantly enriched over input ($mean_{ChIP}/mean_{Input}$, $p \leq 0.05$), (3) significantly enriched over mock IP ($mean_{ChIP}/mean_{mock\ IP}$,

$p \leq 0.05$). To evaluate the statistical stringency of these standards, I randomly sampled 50,000 genome locations and assigned the p-value as the average percentile of the bound sites in 50 independent trials ($p = 0.0066$).

4.7. *In vivo* nucleosome mapping

~250 OD units of cells were collected for each sample. Cells Were crosslinked with 1% formaldehyde for 15 minutes and quenched with 125 mM glycine for 5 minutes at room temperature. Cells were collected by centrifugation, immediately washed with buffer containing 10 mM Tris pH 7.5, 100 mM NaCl, and mechanically lysed with glass beads in lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate). The pellet chromatin fraction was washed twice with MNase reaction buffer (10 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM b-ME, 0.1% Igepal) and digested to primarily mono-nucleosomes with 0.5 – 2 U MNase at 37°C for 30 minutes. Crosslinks Were reversed in the samples and proteins digested by incubation with proteinase K (0.2 mg/ml) at 65 °C for at least 6 hours. DNA was then precipitated with ethanol and digested with 20 ug/ml RNase to remove residual RNA. Mono-nucleosomes were isolated from a 1.5% agarose gel and extracted with a Freeze 'N Squeeze DNA gel extraction column (BioRad) folloId by ethanol precipitation. Sequencing libraries were prepared following the Illumina paired-end protocol.

Libraries were sequenced from both ends with an Illumina genome analyzer II and aligned to the *S. cerevisiae* genome with ELAND. Sequence tags that mapped to more than one genome location and those tags with two ends mapping onto different chromosomes Were removed from this study. I assumed that the center of each sequenced DNA fragment is the nucleosome dyad (Mavrich et al., 2008), and thus aligned the center of each sequenced nucleosomal DNA

fragment on the genome and extended 73 bp on both sides to generate mono-nucleosome coverage. The total nucleosome occupancy across the whole genome was normalized to the same number.

I obtained nucleosome maps for the wild-type strain in high and no Pi conditions (40 minutes after Pi starvation) and for a *cbf1Δ* strain in high Pi conditions. For each sample, 8 – 10 million uniquely aligned nucleosomal DNA sequencing reads were obtained.

Our nucleosome occupancy map is highly correlated with (R = 0.737 at single base pair resolution) the existing genome-wide nucleosome occupancy maps (Kaplan et al., 2008). Considering the difference in yeast culturing conditions between these studies, our map is in consistent with previous published data sets.

4.8. Microarray and data processing

To capture transcription profiles at a particular time during the course of Pi starvation, yeast cells were collected with a cold methanol based quenching method (Pieterse et al., 2006). Cell cultures were added directly into cold methanol (~ -50 °C) with a volume ratio of 2:3, and incubated in an ethanol-dry ice bath for at least 20 minutes. Cells were collected by centrifugation and quickly washed with ice-cold water to remove alcohol, and resuspended in RNAlater solution (Ambion). For each sample, 5×10^7 cells were used to isolate total RNA with the RNase Mini kit (Qiangen), and RNA integrity was analyzed on an agarose gel or with an Agilent bioanalyzer. cDNA was synthesized from 10 μg total RNA with 1:1 ratio of random 10-mers and oligo-dT primers (Operon) and a 2:3 ratio of amino allyl-UTP:dTTP (Sigma), using the Superscript III reverse transcription system (Invitrogen). cDNA was purified with a PCR purification kit (Qiagen) after hydrolyzing RNA. Purified cDNA samples were then labeled with

NHS-ester Cy3 or Cy5 (GE Biosciences). 300 ng Cy3-labeled and 300 ng Cy5-labeled cDNA was competitively hybridized to Agilent 8×15K *S. cerevisiae* two-color expression microarrays (G2509F) in Agilent hybridization buffer for 17 hours at 60 °C. Microarrays Were washed and scanned immediately using an Axon 4000B scanner (Vijayan et al., 2009). The average intensity of the Cy3 and Cy5 fluorescence at each spot was then extracted using GenePix 5.1 software. Lowess and quantile normalization were performed using the MATLAB bioinformatics toolbox before further analysis.

Wild-type (WT) no Pi vs WT high Pi, *cbf1Δ* high Pi vs WT high Pi, *cbf1Δ pho4*Δ high Pi vs WT high Pi, and *tye7Δ rtg3Δ* high Pi vs WT high Pi microarrays Were performed with dye-swaps to eliminate dye labeling bias and analyzed in four biological replicates (Churchill, 2002; Yang and Speed, 2002). *pho80Δ* vs *pho80Δ pho4Δ* and *pho80Δ cbf1Δ* vs *pho80Δ cbf1Δ pho4Δ* Were performed in high Pi conditions and analyzed in two biological replicates. Mutant cycle analysis was constructed with a cyclic comparison so that the expression components could be directly inferred and dye labeling bias would be cancelled in the analysis (Churchill, 2002; Quackenbush, 2002; Yang and Speed, 2002). The mutant cycle was repeated with three biological replicates.

4.9. Fluorescence microscopy

Strains expressing Pho4-GFP were grown in high Pi medium until early/mid- logarithmic phase (OD_{600} 0.3 ~ 0.4), or starved for Pi in Pi-free medium for approximately 1 hour. 1 ml cultures were collected by brief centrifugation and resuspended in ~ 50 ml residual medium. 1 ml of the cell suspension was deposited onto a thin agar pad containing the same medium as the cell culture and fluorescence images were obtained with a Nikon inverted microscope fitted with

an oil-immersion Nikon 40X objective. The exposure time was pre-set to 2.5 s for all images and 5 – 8 separated fields were taken for each sample.

5. References

Adams, C.C., and Workman, J.L. (1995). Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Mol Cell Biol *15*, 1405-1421.

Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome. Nature *446*, 572-576.

Arndt, K.T., Styles, C., and Fink, G.R. (1987). Multiple global regulators control HIS4 transcription in yeast. Science *237*, 874-880.

Audhya, A., and Emr, S.D. (2002). Stt4 PI 4-kinase localizes to the plasma membrane and functions in the Pkc1-mediated MAP kinase cascade. Developmental cell *2*, 593-605.

Audhya, A., Foti, M., and Emr, S.D. (2000). Distinct roles for the yeast phosphatidylinositol 4-kinases, Stt4p and Pik1p, in secretion, cell growth, and organelle membrane dynamics. Mol Biol Cell *11*, 2673-2689.

Auesukaree, C., Homma, T., Tochio, H., Shirakawa, M., Kaneko, Y., and Harashima, S. (2004). Intracellular phosphate serves as a signal for the regulation of the PHO pathway in Saccharomyces cerevisiae. J Biol Chem *279*, 17289-17294.

Ausubel, F.M. (1987). Current protocols in molecular biology (New York: J. Wiley).

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Bahler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., 3rd, Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943-951.

Barbaric, S., Fascher, K.D., and Horz, W. (1992). Activation of the weakly regulated PHO8 promoter in S. cerevisiae: chromatin transition and binding sites for the positive regulatory protein PHO4. Nucleic Acids Res *20*, 1031-1038.

Barbaric, S., Munsterkotter, M., Goding, C., and Horz, W. (1998). Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol Cell Biol *18*, 2629-2639.

Barbaric, S., Munsterkotter, M., Svaren, J., and Horz, W. (1996). The homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast PHO5 promoter. Nucleic Acids Res *24*, 4479-4486.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Blaiseau, P.L., and Thomas, D. (1998). Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA. EMBO J *17*, 6327-6336.

Bulyk, M.L. (2003). Computational prediction of transcription-factor binding site locations. Genome Biol *5*, 201.

Cai, M., and Davis, R.W. (1990). Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. Cell *61*, 437-446.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet *40*, 1300-1306.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. Nat Genet *32 Suppl*, 490-495.

Crespo, J.L., Powers, T., Fowler, B., and Hall, M.N. (2002). The TOR-controlled transcription activators GLN3, RTG1, and RTG3 are regulated in response to intracellular levels of glutamine. Proc Natl Acad Sci U S A *99*, 6784-6789.

Daignan-Fornier, B., and Fink, G.R. (1992). Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. Proc Natl Acad Sci U S A *89*, 6746-6750.

de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., Walther, T.C., and Mann, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature *455*, 1251-1254.

Ellenberger, T., Fass, D., Arnaud, M., and Harrison, S.C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Dev *8*, 970-980.

Ephrussi, A., Church, G.M., Tonegawa, S., and Gilbert, W. (1985). B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science *227*, 134-140.

Ferre-D'Amare, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J *13*, 180-189.

Ferre-D'Amare, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature *363*, 38-45.

Friberg, M.T. (2007). Prediction of transcription factor binding sites using ChIP-chip and phylogenetic footprinting data. Journal of bioinformatics and computational biology *5*, 105-116.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Goke, J., Jung, M., Behrens, S., Chavez, L., O'Keeffe, S., Timmermann, B., Lehrach, H., Adjaye, J., and Vingron, M. (2011). Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. PLoS computational biology *7*, e1002304.

Granek, J.A., and Clarke, N.D. (2005). Explicit equilibrium modeling of transcription-factor binding and gene regulation. Genome Biol *6*, R87.

Graumann, J., Dunipace, L.A., Seol, J.H., McDonald, W.H., Yates, J.R., 3rd, Wold, B.J., and Deshaies, R.J. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. Mol Cell Proteomics *3*, 226-237.

Guthrie, C., and Fink, G.R. (1991). Guide to yeast genetics and molecular biology (San Diego: Academic Press).

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J.*, et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Hattori, N., Kawamoto, H., Fujimoto, S., Kuno, K., and Katsura, Y. (1996). Involvement of transcription factors TCF-1 and GATA-3 in the initiation of the earliest step of T cell development in the thymus. The Journal of experimental medicine *184*, 1137-1147.

Ho, I.C., and Pai, S.Y. (2007). GATA-3 - not just for Th2 cells anymore. Cellular & molecular immunology *4*, 15-29.

Ho, J.W., Bishop, E., Karchenko, P.V., Negre, N., White, K.P., and Park, P.J. (2011). ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC genomics *12*, 134.

Hosoya, T., Kuroha, T., Moriguchi, T., Cummings, D., Maillard, I., Lim, K.C., and Engel, J.D. (2009). GATA-3 is required for early T lineage progenitor development. The Journal of experimental medicine *206*, 2987-3000.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Jia, Y., Rothermel, B., Thornton, J., and Butow, R.A. (1997). A basic helix-loop-helix-leucine zipper transcription complex in yeast functions in a signaling pathway from mitochondria to the nucleus. Mol Cell Biol *17*, 1110-1117.

Kaffman, A., Herskowitz, I., Tjian, R., and O'Shea, E.K. (1994). Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. Science *263*, 1153-1156.

Kaffman, A., Rank, N.M., O'Neill, E.M., Huang, L.S., and O'Shea, E.K. (1998a). The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. Nature *396*, 482-486.

Kaffman, A., Rank, N.M., and O'Shea, E.K. (1998b). Phosphorylation regulates association of the transcription factor Pho4 with its import receptor Pse1/Kap121. Genes Dev *12*, 2673-2683.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kent, N.A., Eibert, S.M., and Mellor, J. (2004). Cbf1p is required for chromatin remodeling at promoter-proximal CACGTG motifs in yeast. J Biol Chem *279*, 27116-27123.

Kent, N.A., Tsang, J.S., Crowther, D.J., and Mellor, J. (1994). Chromatin structure modulation in Saccharomyces cerevisiae by centromere and promoter factor 1. Mol Cell Biol *14*, 5229-5241.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259-272.

Kim, H.D., and O'Shea, E.K. (2008). A quantitative model of transcription factor-activated gene expression. Nat Struct Mol Biol *15*, 1192-1198.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. BMC Mol Biol *10*, 6.

Komeili, A., and O'Shea, E.K. (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science *284*, 977-980.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kuras, L., Cherest, H., Surdin-Kerjan, Y., and Thomas, D. (1996). A heteromeric complex containing the centromere binding factor 1 and two basic leucine zipper factors, Met4 and Met28, mediates the transcription activation of yeast sulfur metabolism. EMBO J *15*, 2519-2529.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246-250.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lee, T.A., Jorgensen, P., Bognar, A.L., Peyraud, C., Thomas, D., and Tyers, M. (2010). Dissection of combinatorial control by the Met4 transcriptional complex. Mol Biol Cell *21*, 456-469.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007a). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Lee, Y.S., Mulugu, S., York, J.D., and O'Shea, E.K. (2007b). Regulation of a cyclin-CDK-CDK inhibitor complex by inositol pyrophosphates. Science *316*, 109-112.

Lenburg, M.E., and O'Shea, E.K. (1996). Signaling phosphate starvation. Trends Biochem Sci *21*, 383-387.

Lohning, C., and Ciriacy, M. (1994). The TYE7 gene of Saccharomyces cerevisiae encodes a putative bHLH-LZ transcription factor required for Ty1-mediated gene expression. Yeast *10*, 1329-1339.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451-459.

MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113.

Maerkl, S.J., and Quake, S.R. (2007). A systems approach to measuring the binding energy landscapes of transcription factors. Science *315*, 233-237.

Magbanua, J.P., Fujisawa, K., Ogawa, N., and Oshima, Y. (1997a). The homeodomain protein Pho2p binds at an A/T-rich segment flanking the binding site of the basic-helix-loop-helix protein Pho4p in the yeast PHO promoters. Yeast *13*, 1299-1308.

Magbanua, J.P., Ogawa, N., Harashima, S., and Oshima, Y. (1997b). The transcriptional activators of the PHO regulon, Pho4p and Pho2p, interact directly with each other and with components of the basal transcription machinery in Saccharomyces cerevisiae. J Biochem *121*, 1182-1189.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol *20*, 429-440.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., *et al.* (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358-362.

McAndrew, P.C., Svaren, J., Martin, S.R., Horz, W., and Goding, C.R. (1998). Requirements for chromatin modulation and transcription activation by the Pho4 acidic activation domain. Mol Cell Biol *18*, 5818-5827.

Mellor, J., Rathjen, J., Jiang, W., Barnes, C.A., and Dowell, S.J. (1991). DNA binding of CPF1 is required for optimal centromere function but not for maintaining methionine prototrophy in yeast. Nucleic Acids Res *19*, 2961-2969.

Miele, A., Bystricky, K., and Dekker, J. (2009). Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. PLoS genetics *5*, e1000478.

Miller, J.A., and Widom, J. (2003). Collaborative competition mechanism for gene activation in vivo. Mol Cell Biol *23*, 1623-1632.

Mitchell, P.J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371-378.

Morozov, A.V., and Siggia, E.D. (2007). Connecting protein structure with predictions of regulatory sites. Proc Natl Acad Sci U S A *104*, 7068-7073.

Munsterkotter, M., Barbaric, S., and Horz, W. (2000). Transcriptional regulation of the yeast PHO8 promoter in comparison to the coregulated PHO5 promoter. J Biol Chem *275*, 22678-22685.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nishi, K., Park, C.S., Pepper, A.E., Eichinger, G., Innis, M.A., and Holland, M.J. (1995). The GCR1 requirement for yeast glycolytic gene expression is suppressed by dominant mutations in the SGC1 gene, which encodes a novel basic-helix-loop-helix protein. Mol Cell Biol *15*, 2646-2653.

Nishizawa, M., Komai, T., Katou, Y., Shirahige, K., Ito, T., and Toh, E.A. (2008). Nutrient-regulated antisense and intragenic RNAs modulate a signal transduction pathway in yeast. PLoS Biol *6*, 2817-2830.

O'Neill, E.M., Kaffman, A., Jolly, E.R., and O'Shea, E.K. (1996). Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science *271*, 209-212.

Ogawa, N., DeRisi, J., and Brown, P.O. (2000). New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. Mol Biol Cell *11*, 4309-4321.

Ogawa, N., Saitoh, H., Miura, K., Magbanua, J.P., Bun-ya, M., Harashima, S., and Oshima, Y. (1995). Structure and distribution of specific cis-elements for transcriptional regulation of PHO84 in Saccharomyces cerevisiae. Molecular & general genetics : MGG *249*, 406-416.

Oshima, Y. (1997). The phosphatase system in Saccharomyces cerevisiae. Genes Genet Syst *72*, 323-334.

Peng, S., Alekseyenko, A.A., Larschan, E., Kuroda, M.I., and Park, P.J. (2007). Normalization and experimental design for ChIP-chip data. BMC Bioinformatics *8*, 219.

Pieterse, B., Jellema, R.H., and van der Werf, M.J. (2006). Quenching of microbial samples for increased reliability of microarray data. J Microbiol Methods *64*, 207-216.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat Genet *32 Suppl*, 496-501.

Robinson, K.A., and Lopes, J.M. (2000). SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res *28*, 1499-1505.

Rothermel, B.A., Thornton, J.L., and Butow, R.A. (1997). Rtg3p, a basic helix-loop-helix/leucine zipper protein that functions in mitochondrial-induced changes in gene expression, contains independent activation domains. J Biol Chem *272*, 19801-19807.

Rouillon, A., Barbey, R., Patton, E.E., Tyers, M., and Thomas, D. (2000). Feedback-regulated degradation of the transcriptional activator Met4 is triggered by the SCF(Met30)complex. EMBO J *19*, 282-294.

Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol *27*, 66-75.

Sato, T., Lopez, M.C., Sugioka, S., Jigami, Y., Baker, H.V., and Uemura, H. (1999). The E-box DNA binding protein Sgc1p suppresses the gcr2 mutation, which is involved in transcriptional activation of glycolytic genes in Saccharomyces cerevisiae. FEBS letters *463*, 307-311.

Schneider, K.R., Smith, R.L., and O'Shea, E.K. (1994). Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science *266*, 122-126.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat Rev Mol Cell Biol *2*, 827-837.

Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997). Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. EMBO J *16*, 4689-4697.

Springer, M., Wykoff, D.D., Miller, N., and O'Shea, E.K. (2003). Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. PLoS Biol *1*, E28.

Stormo, G.D., and Fields, D.S. (1998). Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci *23*, 109-113.

Stormo, G.D., and Zhao, Y. (2010). Determining the specificity of protein-DNA interactions. Nat Rev Genet *11*, 751-760.

Svaren, J., and Horz, W. (1997). Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. Trends Biochem Sci *22*, 93-97.

Svaren, J., Schmitz, J., and Horz, W. (1994). The transactivation domain of Pho4 is required for nucleosome disruption at the PHO5 promoter. EMBO J *13*, 4856-4862.

Thomas, D., Jacquemin, I., and Surdin-Kerjan, Y. (1992). MET4, a leucine zipper protein, and centromere-binding factor 1 are both required for transcriptional activation of sulfur metabolism in Saccharomyces cerevisiae. Mol Cell Biol *12*, 1719-1727.

Thomas, M.R., and O'Shea, E.K. (2005). An intracellular phosphate buffer filters transient fluctuations in extracellular phosphate levels. Proc Natl Acad Sci U S A *102*, 9565-9570.

Tice-Baldwin, K., Fink, G.R., and Arndt, K.T. (1989). BAS1 has a Myb motif and activates HIS4 transcription only in combination with BAS2. Science *246*, 931-935.

van Werven, F.J., and Timmers, H.T. (2006). The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. Nucleic Acids Res *34*, e33.

Vashee, S., Melcher, K., Ding, W.V., Johnston, S.A., and Kodadek, T. (1998). Evidence for two modes of cooperative DNA binding in vivo that do not involve direct protein-protein interactions. Curr Biol *8*, 452-458.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Venter, U., Svaren, J., Schmitz, J., Schmid, A., and Horz, W. (1994). A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. EMBO J *13*, 4848-4855.

Vijayan, V., Zuzow, R., and O'Shea, E.K. (2009). Oscillations in supercoiling drive circadian gene expression in cyanobacteria. Proc Natl Acad Sci U S A *106*, 22564-22568.

Vogel, K., Horz, W., and Hinnen, A. (1989). The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. Mol Cell Biol *9*, 2050-2057.

Waldminghaus, T., and Skarstad, K. (2010). ChIP on Chip: surprising results are often artifacts. BMC genomics *11*, 414.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet *11*, 426-437.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J *29*, 2147-2160.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res *20*, 90-100.

Yang, Y.H., and Speed, T. (2002). Design issues for cDNA microarray experiments. Nat Rev Genet *3*, 579-588.

Yoshida, S., Ohya, Y., Goebl, M., Nakano, A., and Anraku, Y. (1994). A novel gene, STT4, encodes a phosphatidylinositol 4-kinase in the PKC1 protein kinase pathway of Saccharomyces cerevisiae. J Biol Chem *269*, 1166-1172.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940-954.

Zhao, G., Ihuegbu, N., Lee, M., Schriefer, L., Wang, T., and Stormo, G.D. (2012). Conserved Motifs and Prediction of Regulatory Modules in Caenorhabditis elegans. G3 (Bethesda) *2*, 469-481.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M.*, et al.* (2009a). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M.*, et al.* (2009b). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res *19*, 556-566.

Chapter III

Conclusions

and

Future Directions

1. Conclusions

Sequence-specific transcriptional factors recognize short *cis* elements in promoter regions to regulate gene transcription (Farnham, 2009; Jacob and Monod, 1961). However, these elements exist tens of times more frequently than the number of *in vivo* binding events. How can transcription factors bind to and regulate a specific set of genes when many other genes carry the same binding sites? Moreover, many recent studies examining the binding of transcription factors at the whole genome scale presented the surprising observation that only a fraction of the binding events are functional, questioning the connection between transcription factor binding and its subsequent influence on gene expression. What are the underlying mechanisms to bridge these critical elements – from sequence motif to transcription factor binding and from binding to transcription function? Beyond the regulation of a single transcription factor, transcription factors with DNA binding domain belonging to the same structural family commonly recognize highly similar DNA sequences, yet they regulate diverse biological processes from budding yeast to human (Badis et al., 2009; Berger et al., 2008; Robinson and Lopes, 2000; Wei et al., 2010). How do these sequence-specific transcription factors with overlapping binding specificity faithfully and distinctly regulate transcription responses to diverse signaling pathways?

The phosphate-responsive signaling pathway in *Saccharomyces cerevisiae* provides us with a convenient system to address the above questions. The signaling processes that directly regulate the transcriptional responses of PHO pathway have been extensively mapped (Huang and O'Shea, 2005); the transcriptional activation is primarily regulated by the transcription factor Pho4, whose DNA binding affinity has been carefully defined (Badis et al., 2008; Maerkl and Quake, 2007; Zhu et al., 2009); transcription factors that contain the similar DNA binding domain as Pho4 are known and they regulate phospholipid biosnythesis, sulfur metabolic

pathways, retrograde responses, expression of glycolytic genes and filamentous growth in the budding yeast (Robinson and Lopes, 2000).

I used a genome-wide approach to systematically dissect the determinants of Pho4 binding and regulation, and investigated whether these determinants can explain transcriptional specificity. I provide evidence that the intrinsic DNA binding affinity of Pho4 is necessary to dictate Pho4 binding *in vivo*. Among all the sequences that have high affinity for Pho4, competition from chromatin and competition from another transcription factor, Cbf1, which recognizes Pho4 consensus binding motifs, determine the selective binding of Pho4 at only a small fraction of these sites.

Not all the binding events are functional even if they are located close to the transcription start site of a gene. A cooperative interaction between Pho2 and Pho4 determines the outcome of the Pho4 binding events – the ability to activate gene transcription. This cooperative interaction is a result of cooperative binding between Pho2 and Pho4, and sequence motif analysis suggests that the spatial arrangement between the binding sequences of Pho2 and Pho4 promotes their cooperative binding at the functional binding sites. This cooperative interaction provides an additional layer of selection for the specificity of PHO regulation.

Pho4 regulates the phosphate starvation response and Cbf1 regulates the sulfur metabolic pathways (MET pathways). Both the PHO genes and the MET genes contain the core 'CACGTG' binding sequences at their promoters. How does each of these factors define their specific regulon and how does the competition between them influence their pathway specific regulation? The competition between Pho4 and Cbf1 is specified by their differential preferences at a single base flanking the consensus binding sequence – the sites with a 'T' base at the 3' of

the ‘CACGTG’ sites are most favored by Cbf1 so that binding of Cbf1 at these sites blocks Pho4. Coincidently, the MET genes carry the ‘TCACGTG’ sequences at their promoter and the PHO genes carry the alternatives. I also discovered novel regulatory functions for Cbf1 competition: it raises the threshold for transcriptional activation by Pho4 to ensure the pathway is turned off in phosphate replete conditions, and it prevents Pho4 activation of genes outside the phosphate regulon during phosphate starvation. Interestingly, Pho4 is able to induce the transcription of several MET genes when Cbf1 is removed from the nucleus, highlighting the important influence of transcription factor competition in defining the pathway specific regulation.

Combining all this information, I was able to develop an equilibrium model to accurately predict the binding pattern of Pho4 at its consensus binding sites as well as the functionality of these binding events. The affinity of the binding sequences, the accessibility to Pho4, the competition with Cbf1, all contribute to the accuracy and predictive power of the model, suggesting critical roles of both the intrinsic and extrinsic determinants of transcription factor binding.

Overall, my work revealed the mechanisms for the specific transcriptional regulation of the phosphate-responsive signaling pathway in *S. cerevisiae*, and advanced our understanding in how the specific regulation by transcription factors is achieved in eukaryotic genome. I also identified novel regulatory functions for the competition between two transcription factors with similar DNA binding specificity, providing a plausible explanation for how factors recognizing similar binding sequences could regulate diverse and distinct transcriptional responses. The accurate prediction from the computational model strongly argued the importance of incorporating *trans* influences into the construction of a functional genomic regulatory network. Although the budding yeast is one of the simplest eukaryotic model organisms, the identified

determinants are generally applicable to other eukaryotic systems. Finally, this work provides a road map for step-wise dissecting the determinants of transcriptional specificity in complex regulatory structures.

2. Future directions

2.1. Modular regulation in transcription networks

I showed in the previous chapter that transcriptional activation by Pho4 is highly dependent on the cooperativity between Pho4 and Pho2 in wild type. I also showed that in the *cbf1*Δ strain, Pho4 is able to activate some of the genes that escaped Cbf1 competition in the wild type strain. These observations immediately raise two questions: is the activation of these newly Pho4-regulated genes in the *cbf1*Δ strain also dependent on Pho2? Is the stringent requirement of Pho2 cooperativity in Pho4-dependent gene activation influenced by Cbf1 competition?

I performed mutant cycle experiments to quantify the regulatory interactions between Pho2 and Pho4, in the conditions with and without the competition from Cbf1 (Figure 47). These experiments were performed in the *pho80*Δ background since *CBF1* deletion causes a delay in Pho4 nuclear translocation. Briefly, the genes activated by Pho4 in the absence of Cbf1 are still dependent on the cooperativity between Pho2 and Pho4 (Figure 48); also, Cbf1 competition does not seem to dramatically change the regulatory interactions between Pho2 and Pho4 in activating the PHO gene expression (Table 7). These results demonstrate that the cooperativity between Pho2 and Pho4 is independent of the competition from Cbf1, suggesting that these two influences could act modularly in determining Pho4 binding and regulation. It raises new questions about the transcriptional regulatory network: are the modular influences for transcription factor binding and regulation a general theme for eukaryotic transcriptional control? Competition and cooperativity exert opposite influence on transcriptional regulation. What are the benefits of evolving a network such that the transcriptional control has to be specified by opposing modules and how do they evolve? How would these modules affect the decoding of

promoter information at a single cell level as well as at a population level? Answering these questions would help to examine the regulatory interactions among transcription factors and advance our knowledge in designing transcriptional regulatory network with programmed regulation.

Table 7. Summary of the epistasis expression analysis in both the *pho80Δ* and the *pho80Δcbf1Δ* strains.

The number of genes with significant Pho2, Pho4 and Cooperative components is listed in the table below.

Strain background	Expression components		
	Pho2	Pho4	Cooperative
pho80Δ	6	3	70
pho80Δcbf1Δ	0	7	78

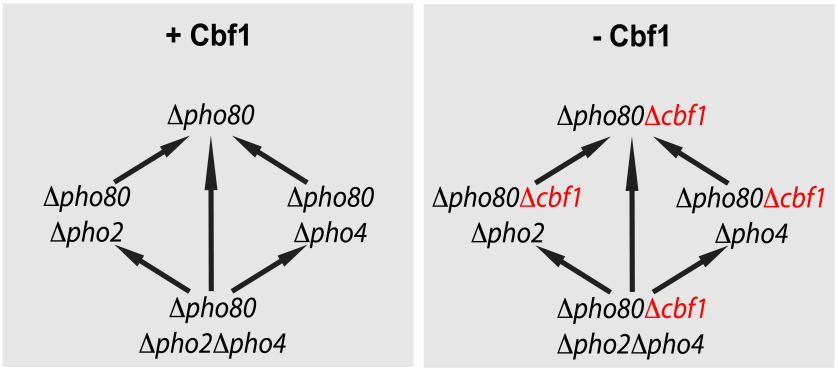


Figure 47. Mutant cycle analyses with and without Cbf1 competition.

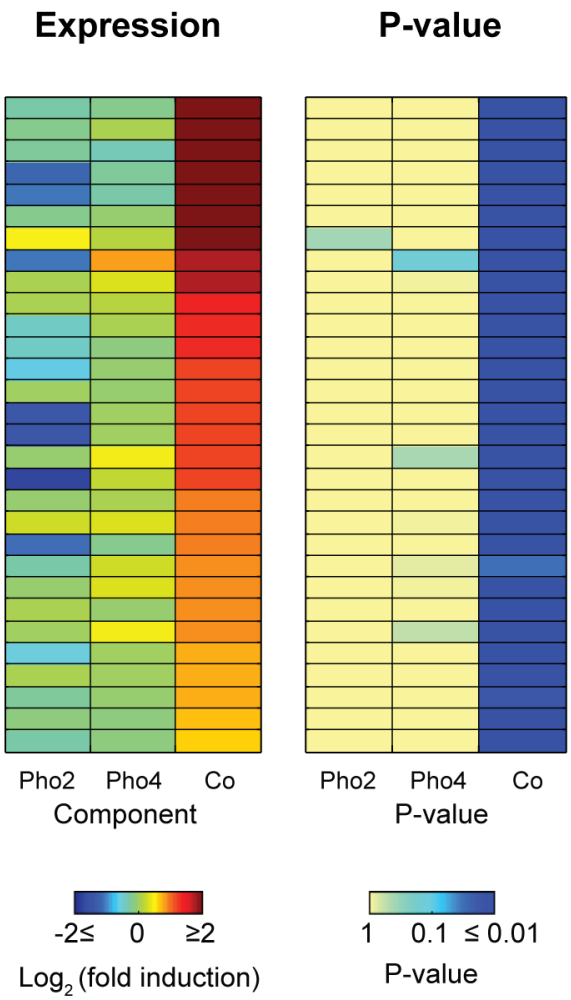


Figure 48. Epistasis expression analysis for the genes that show an increase in Pho4-dependent activation in the *cbf1Δ* strain.

Heat map showing the best fit of expression components (left columns) and their statistical significance (right columns) for genes that show an increase in Pho4-dependent activation comparing the *pho80Δ* and the *pho80Δcbf1Δ* strains.

2.2. Cooperative interactions and their role in the evolution of pathway specific responses

Ascomycota fungi provide a unique species group to study the evolution of regulatory networks (Thompson and Regev, 2009; Wohlbach et al., 2009). A pioneering study by Gasch et. al showed the conservation of transcription factors and their *cis* regulatory elements across ascomycete species (Gasch et al., 2004). Since then, many studies have been focusing on co-evolution of the transcription factors with the change of *cis* regulatory sequences (Gasch et al., 2004), substitution of factors within the regulatory network, and gain and loss of gene regulation across multiple species (Li and Johnson, 2010). However, not much attention is given to the conserved regulatory structure, such as the cooperative regulation between transcription factors. What are the functions of cooperativity in the course of evolution and why would the interaction between two or more transcription factors be conserved across species separated by millions of years? A simple is answer is that the cooperativity is required for activating gene transcription. If this is the case, however, why not choose the alternative where binding of a single transcription factor is sufficient to induce gene transcription?

The regulation of the PHO pathway may present a suitable platform to examine the roles of cooperativity in the evolution of phosphate starvation responses. The cooperativity between Pho2 and Pho4 is necessary for the induction of the PHO genes in *S. cerevisiae*, whereas Pho2 is dispensable in the regulation of PHO genes in *Candida glabrata* (Kerwin and Wykoff, 2009) and *Candida albicans* (Noble et al., 2010), two pathogens that invade human with a compromised immune system (Fidel et al., 1999). Comparing the regulatory interactions between the orthologs of Pho2 and Pho4 in different species might shed light on the reasons for having a conserved cooperativity. Preliminary experiments indicate that Pho4 from *C. glagrata* can activate the PHO genes in the context of *S. cerevisiae* genome, suggesting that the cooperativity

with Pho2 is not a conserved requirement for transcriptional activation. Many questions await answers. How would the cooperativity between Pho2 and Pho4 evolve and how necessary is it a requirement for PHO gene activation across ascomycete species? What are the advantages of gaining or losing the cooperativity in these species? What are the molecular mechanisms of the change of this cooperativity? And can we provide evidence to connect the changes in cooperativity with evolutionary fitness?

2.3. Transcription factor binding, genome-wide chromatin remodeling, and gene regulation

Genome-wide chromatin remodeling in response to environmental stimuli has been studied for conditions such as heat shock and alternative carbon resources, in which the majority of the yeast genome undergoes a significant expression change (Kaplan et al., 2009; Shivaswamy et al., 2008; Zawadzki et al., 2009). However, the relationships between nucleosome remodeling and gene regulation is still unclear (Zawadzki et al., 2009). Specific binding of transcription factors, the subsequent recruitment of chromatin remodelers, secondary responses derived from the primary regulation and variations embedded in the genome-wide approaches may all contribute to this discrepancy. Examining the relationships between transcriptional regulation and nucleosome remodeling in a specific signaling pathway, where the global expression programs and nucleosome profiles do not change significantly, would help to establish a clear picture for the connections between transcription factor, nucleosome remodeling, and gene expression. With the knowledge of Pho4 and its regulation, the genes responsive to phosphate starvation, and the components involved in the steps of chromatin remodeling (Barbaric et al., 2007; Gregory et al., 1999; Huang and O'Shea, 2005), the phosphate starvation response emerges as a great system to examine these questions and a platform to explore the mechanisms underlying these relationships. Preliminary results indicate that the chromatin remodeling at the promoter and the coding region of genes show specific patterns for the genes activated and repressed by phosphate starvation (Figure 49).

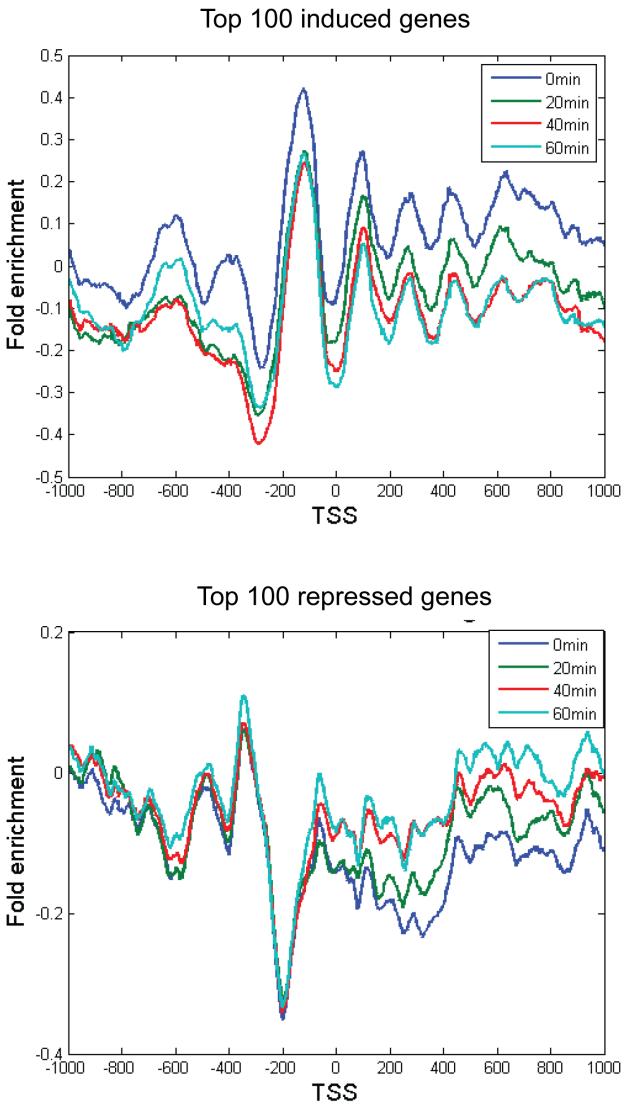


Figure 49. Differential pattern of nucleosome remodeling at promoter and coding regions.

The nucleosome occupancy maps of 1000 bp up- and down- stream of transcription start site (TSS) for the most induced (top 100) and most repressed genes are aligned at their TSS and averaged respectively. The average occupancies are further normalized to the genome average and the relative enrichment of nucleosome occupancy along the promoter and coding sequence is shown above. Clear nucleosome-deplete is present at both the promoter and coding region of the induced genes; while only appearance of nucleosomes is observed in the coding region for the repressed genes.

Transcription factors usually recruit chromatin remodelers or components of the transcription initiation complex to initiate the transcriptional activation program (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997). In *S. cerevisiae*, transcription start sites of repressed genes are usually located within a well-positioned nucleosome when these genes are not induced (Lee et al., 2007; Yuan et al., 2005). It is thought that this nucleosome prevents the assembly of the transcription initiate complex and keeps these genes transcriptionally inactive; induction of those genes requires remodeling of the promoter nucleosome after transcription factor binding to expose their transcription start sites. The initiation of this process has been well studied at a few gene promoters (Weake and Workman, 2010), but spatial range of chromatin remodeling and its implications in gene expression is not well explored. Does chromatin remodeling happen digitally on the order of a single nucleosome, or locally at the scale of an individual gene, or over a continuous genome location? What are necessary landmarks to define the geographic range of these types of regulation? After chromatin remodelers recruited by transcription factors, does chromatin remodeling only happen in the direction of assembling transcription initiation complex? If yes, would additional epigenetic markers be needed to ensure the unidirectional remodeling of nucleosomes? If no, since yeast genome is compact, how would neighboring genes be influenced by bi-directional chromatin remodeling? Answering these questions would help us understanding the specificity of transcriptional regulation, not only at specific loci, but also in the context of its genomic environment.

Substantial progress has been made in our understanding of the mechanisms of transcription since the discovery of transcription factors over 30 years ago (Tjian et al., 1978). The powerful genetic, biochemical and structural analyses have led to vivid pictures of how transcription factors activate the expression of a gene. However, we still do not possess the courage to claim that we understand fully for some of the most basic questions. For example, how would a transcription factor specifically select a fraction of the genes to regulate in the genome of mouse or human? And to put this question into a biology context, why? With the development of high throughput sequencing and other novel biochemical approaches, the field of eukaryotic transcription is entering into a new era with unprecedented information, and these questions that would have not been possible to answer before await for their challengers.

3. References

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X.*, et al.* (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720-1723.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L.*, et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Barbaric, S., Luckenbach, T., Schmid, A., Blaschke, D., Horz, W., and Korber, P. (2007). Redundancy of chromatin remodeling pathways for the induction of the yeast PHO5 promoter in vivo. J Biol Chem *282*, 27610-27621.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. Nat Rev Genet *10*, 605-616.

Fidel, P.L., Jr., Vazquez, J.A., and Sobel, J.D. (1999). Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans. Clinical microbiology reviews *12*, 80-96.

Gasch, A.P., Moses, A.M., Chiang, D.Y., Fraser, H.B., Berardini, M., and Eisen, M.B. (2004). Conservation and evolution of cis-regulatory systems in ascomycete fungi. PLoS Biol *2*, e398.

Gregory, P.D., Schmid, A., Zavari, M., Munsterkotter, M., and Horz, W. (1999). Chromatin remodelling at the PHO8 promoter requires SWI-SNF and SAGA at a step subsequent to activator binding. EMBO J *18*, 6407-6414.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev *17*, 1309-1320.

Huang, S., and O'Shea, E.K. (2005). A systematic high-throughput screen of a yeast deletion collection for mutants defective in PHO5 regulation. Genetics *169*, 1859-1871.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362-366.

Kerwin, C.L., and Wykoff, D.D. (2009). Candida glabrata PHO4 is necessary and sufficient for Pho2-independent transcription of phosphate starvation genes. Genetics *182*, 471-479.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Li, H., and Johnson, A.D. (2010). Evolution of transcription networks--lessons from yeasts. Curr Biol *20*, R746-753.

Maerkl, S.J., and Quake, S.R. (2007). A systems approach to measuring the binding energy landscapes of transcription factors. Science *315*, 233-237.

Noble, S.M., French, S., Kohn, L.A., Chen, V., and Johnson, A.D. (2010). Systematic screens of a Candida albicans homozygous deletion library decouple morphogenetic switching and pathogenicity. Nat Genet *42*, 590-598.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Robinson, K.A., and Lopes, J.M. (2000). SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res *28*, 1499-1505.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS Biol *6*, e65.

Thompson, D.A., and Regev, A. (2009). Fungal regulatory evolution: cis and trans in the balance. FEBS letters *583*, 3959-3965.

Tjian, R., Fey, G., and Graessmann, A. (1978). Biological activity of purified simian virus 40 T antigen proteins. Proc Natl Acad Sci U S A *75*, 1279-1283.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat Rev Genet *11*, 426-437.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J *29*, 2147-2160.

Wohlbach, D.J., Thompson, D.A., Gasch, A.P., and Regev, A. (2009). From elements to modules: regulatory evolution in Ascomycota fungi. Current opinion in genetics & development *19*, 571-578.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626-630.

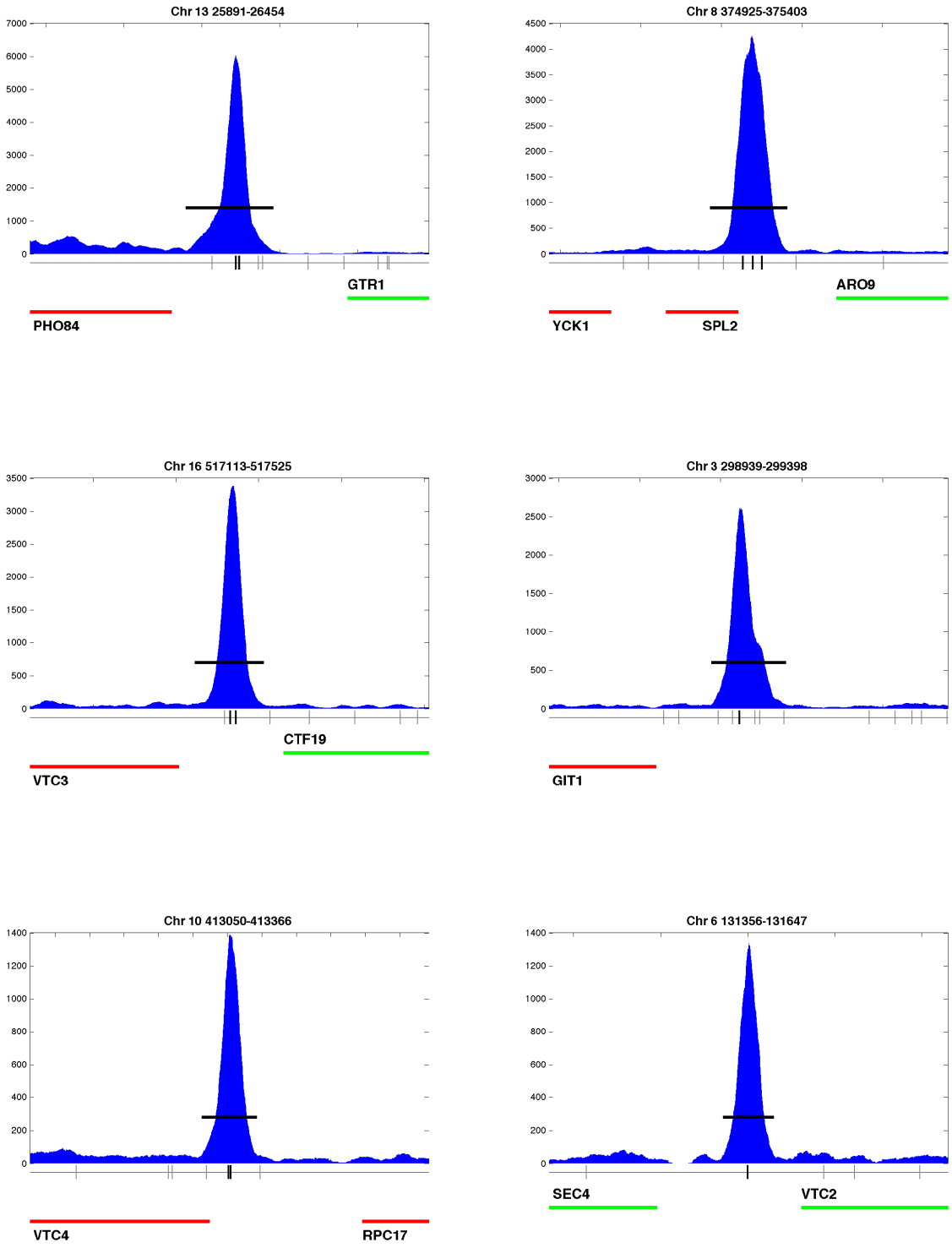
Zawadzki, K.A., Morozov, A.V., and Broach, J.R. (2009). Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in Saccharomyces cerevisiae. Mol Biol Cell *20*, 3503-3513.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M.*, et al.* (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res *19*, 556-566.

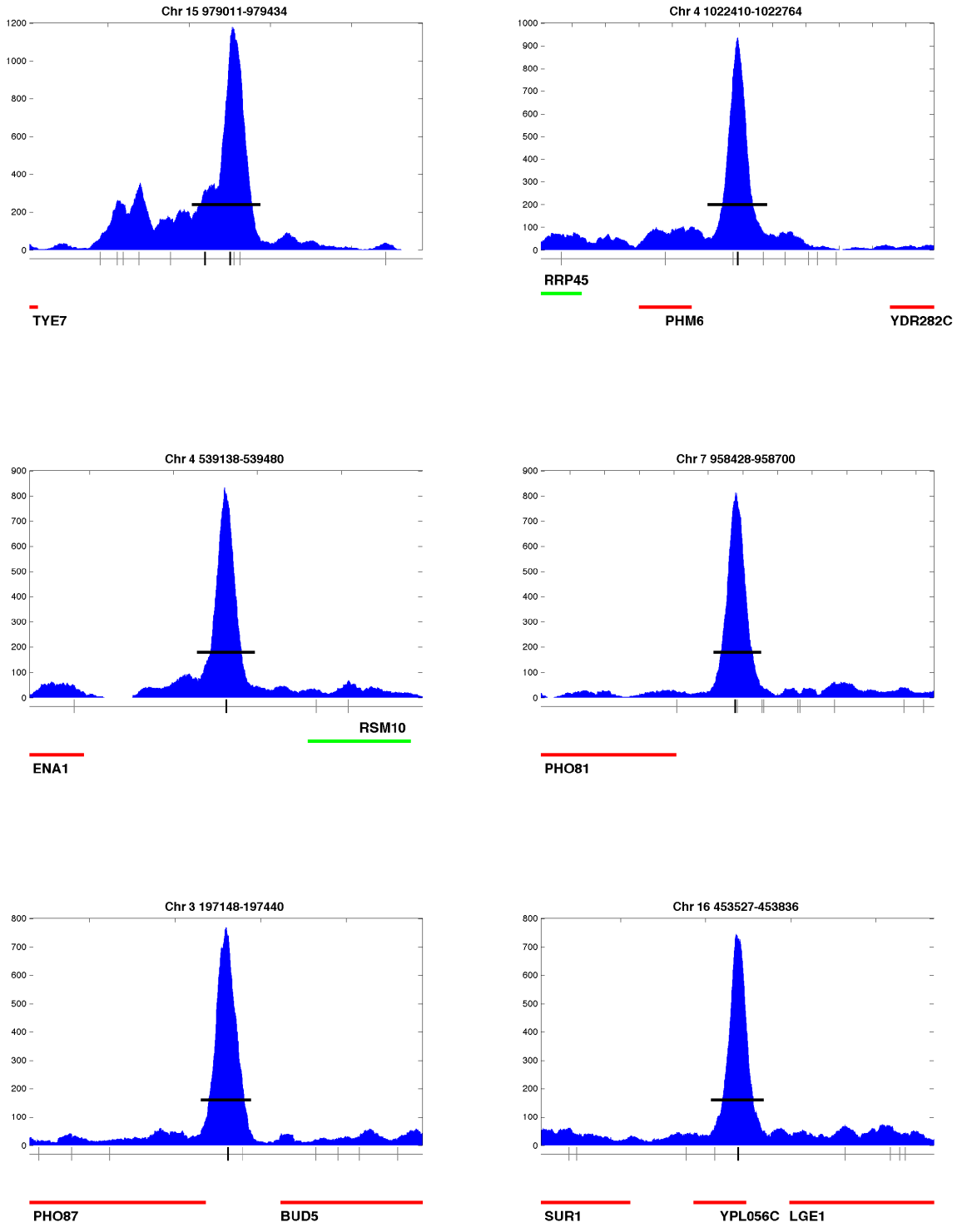
Appendix

Appendix 1. Pho4-bound regions containing high affinity Pho4 binding sites.

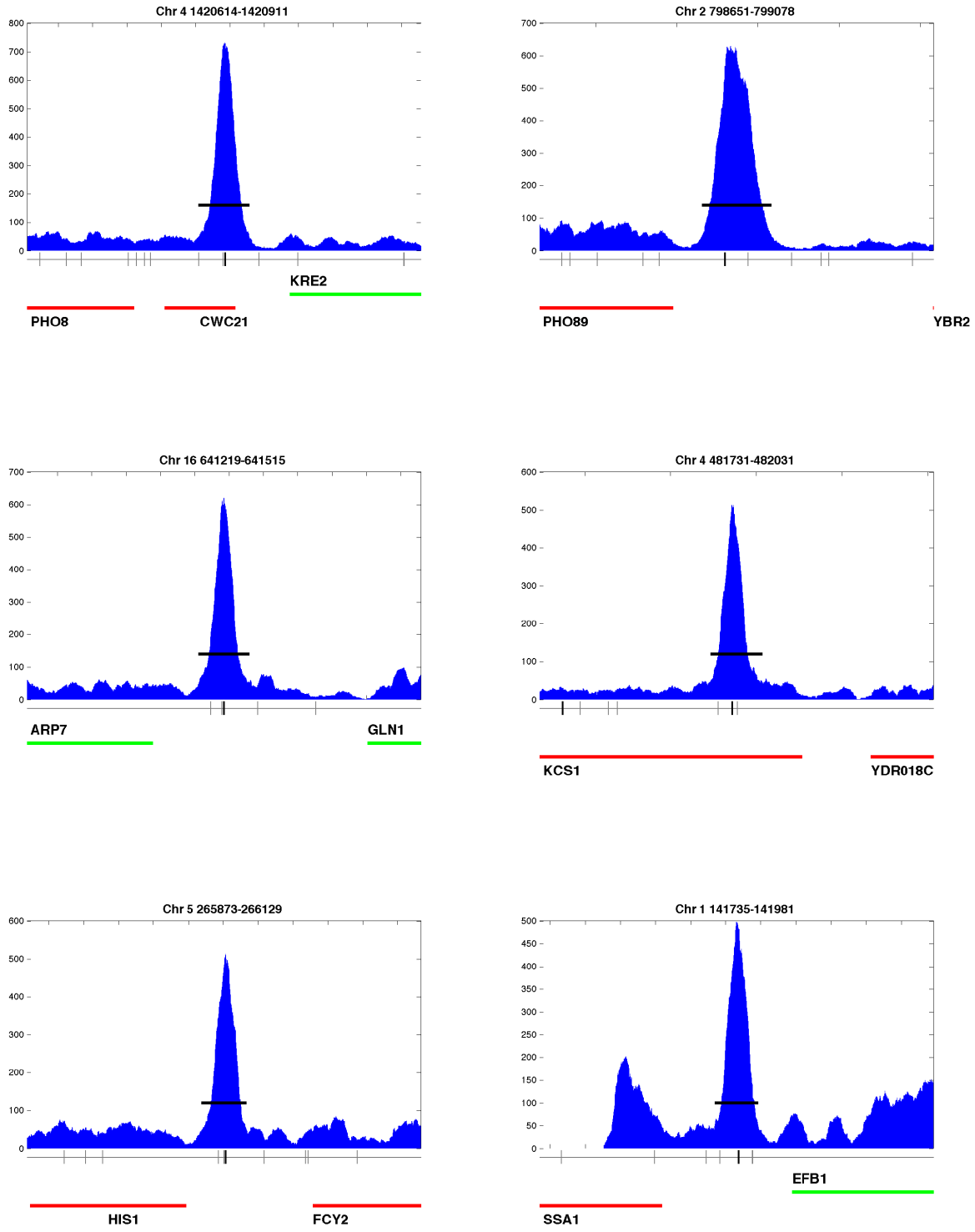
The plots show the Pho4 binding occupancy in no Pi conditions for the genomic region indicated above the plots. Black horizontal lines denote the regions that are determined as enriched of Pho4 binding ($p \leq 0.05$ for enrichment over genomic input and mock IP, enrichment $\geq$ 2-fold). The black vertical lines below the plots indicate the positions of high affinity binding sites, and the gray vertical lines indicate the positions of low affinity binding sites. The open reading frames on the forward (Watson) strand is colored in green and the ORFs on the reverse (Crick) strand is colored in red. The plots show 66 regions in total.



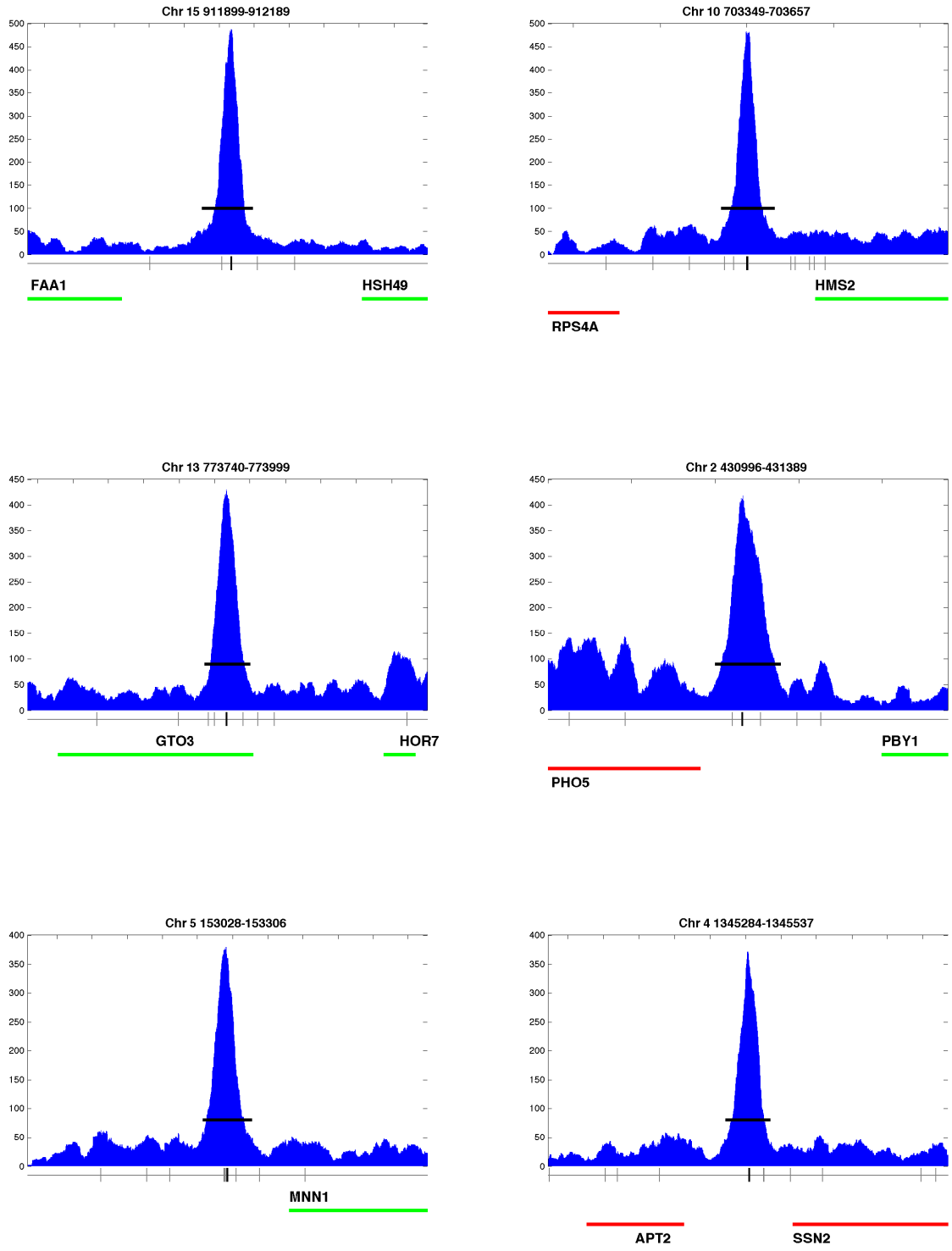
Supplemental Figure 1. Pho4-bound regions containing high affinity Pho4 binding sites



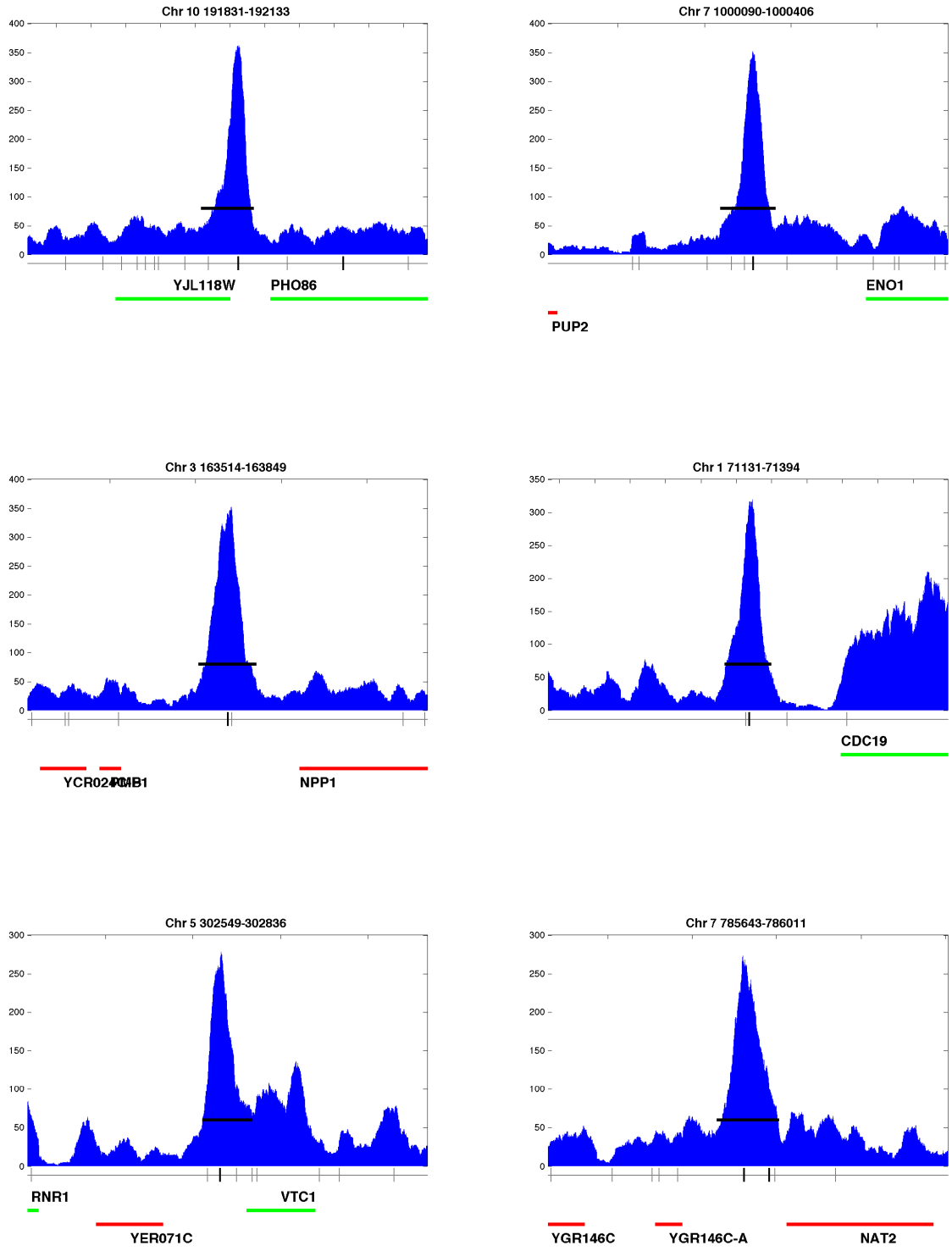
Supplemental Figure 1 (continued).



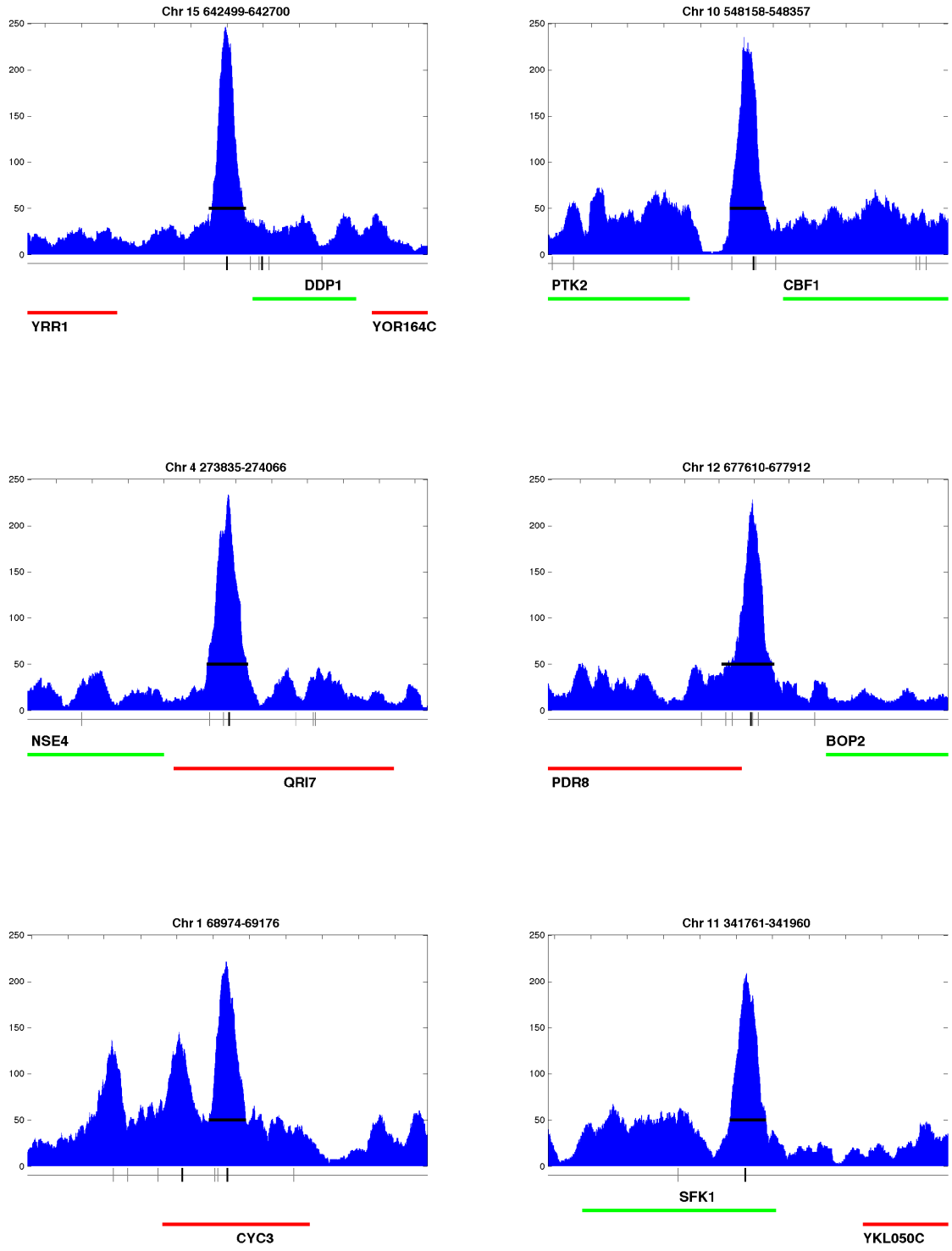
Supplemental Figure 1 (continued).



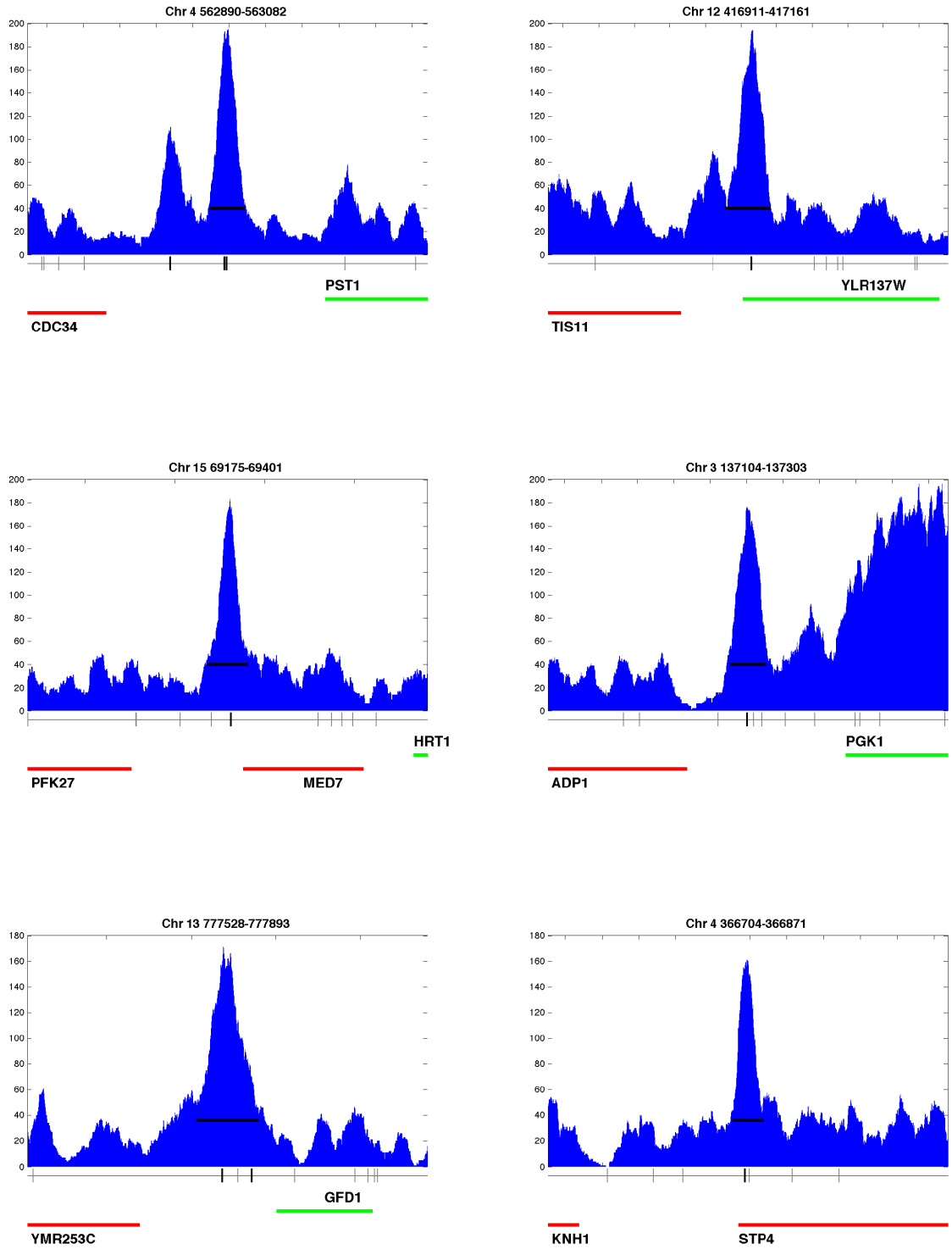
Supplemental Figure 1 (continued).



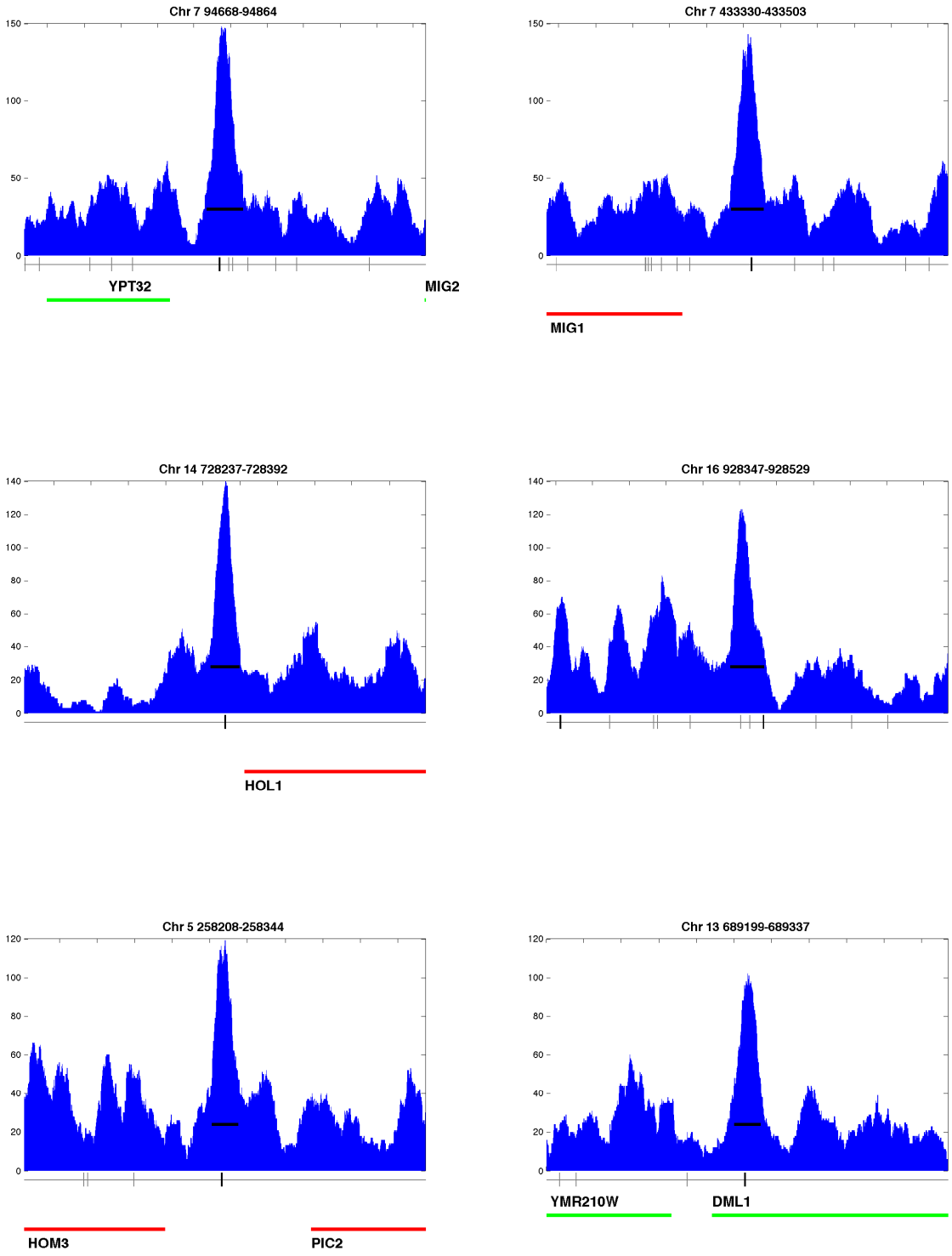
Supplemental Figure 1 (continued).



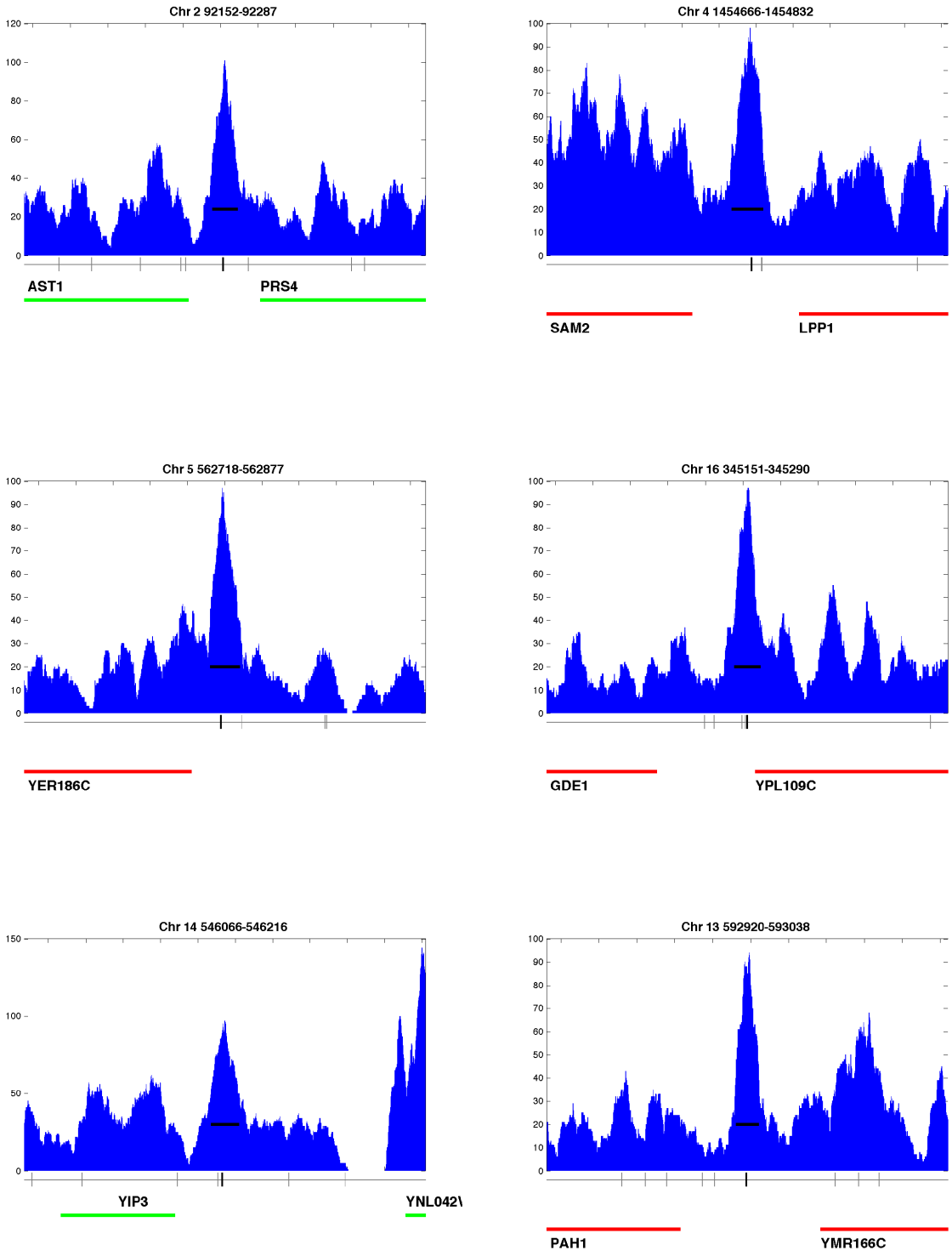
Supplemental Figure 1 (continued).



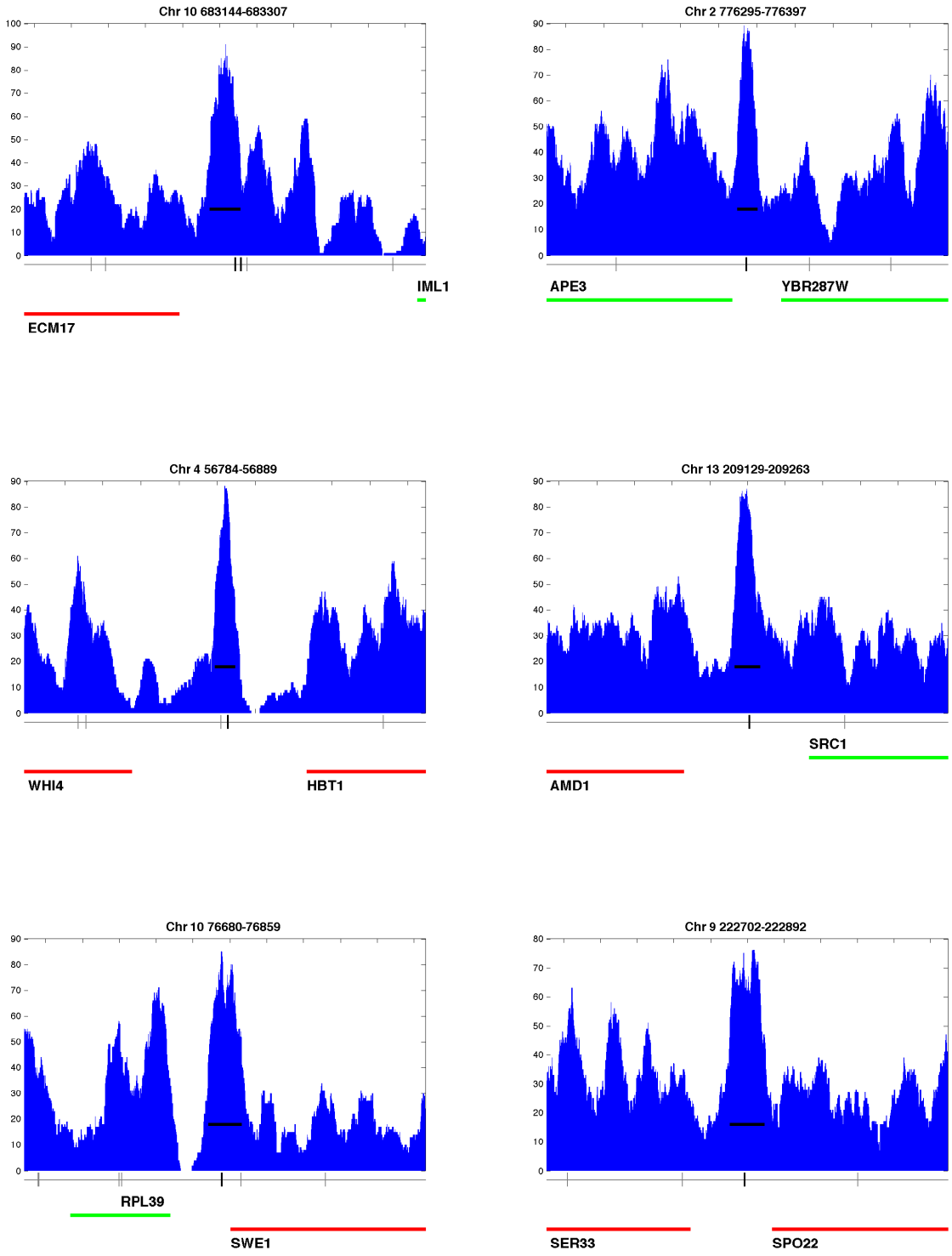
Supplemental Figure 1 (continued).



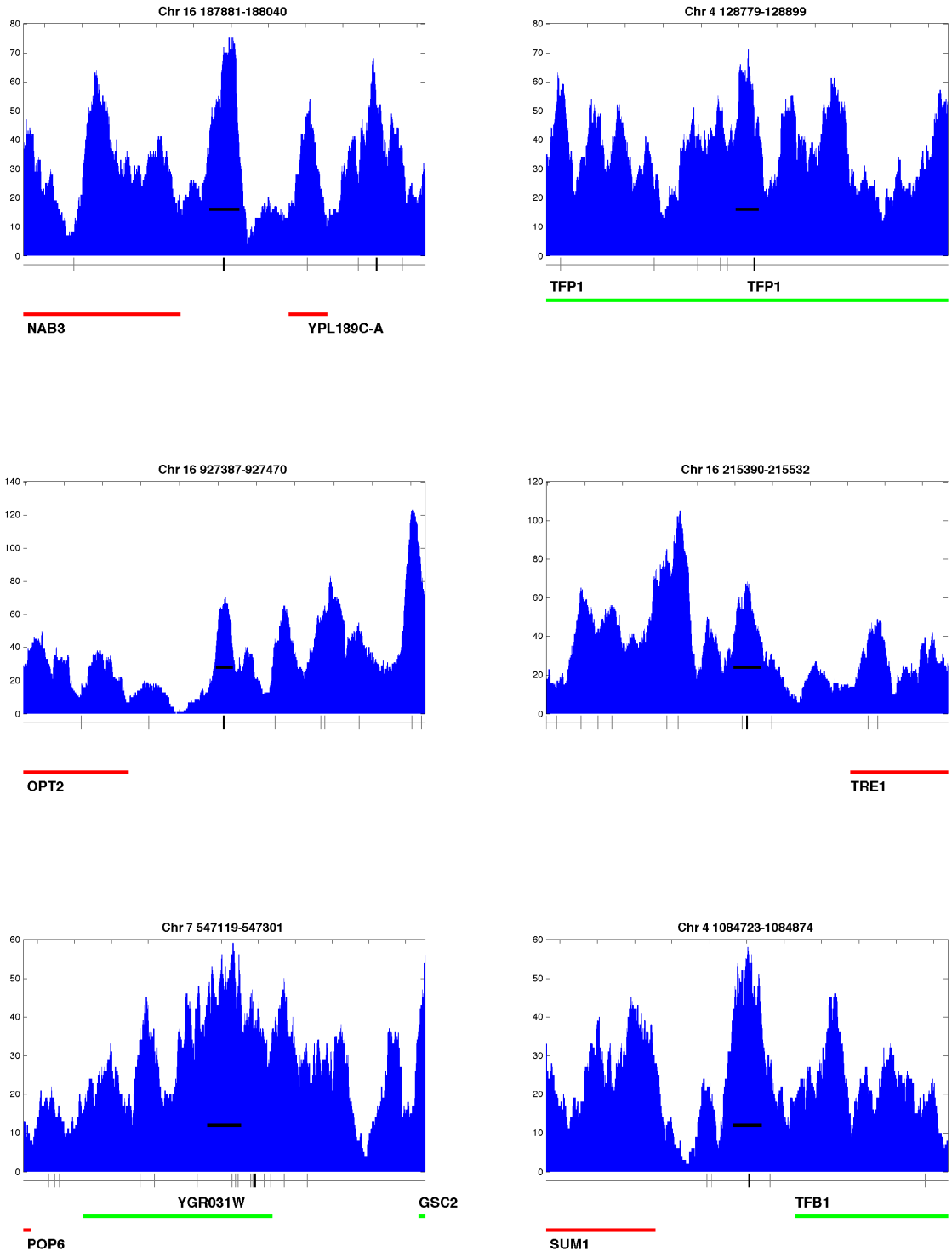
Supplemental Figure 1 (continued).



Supplemental Figure 1 (continued).



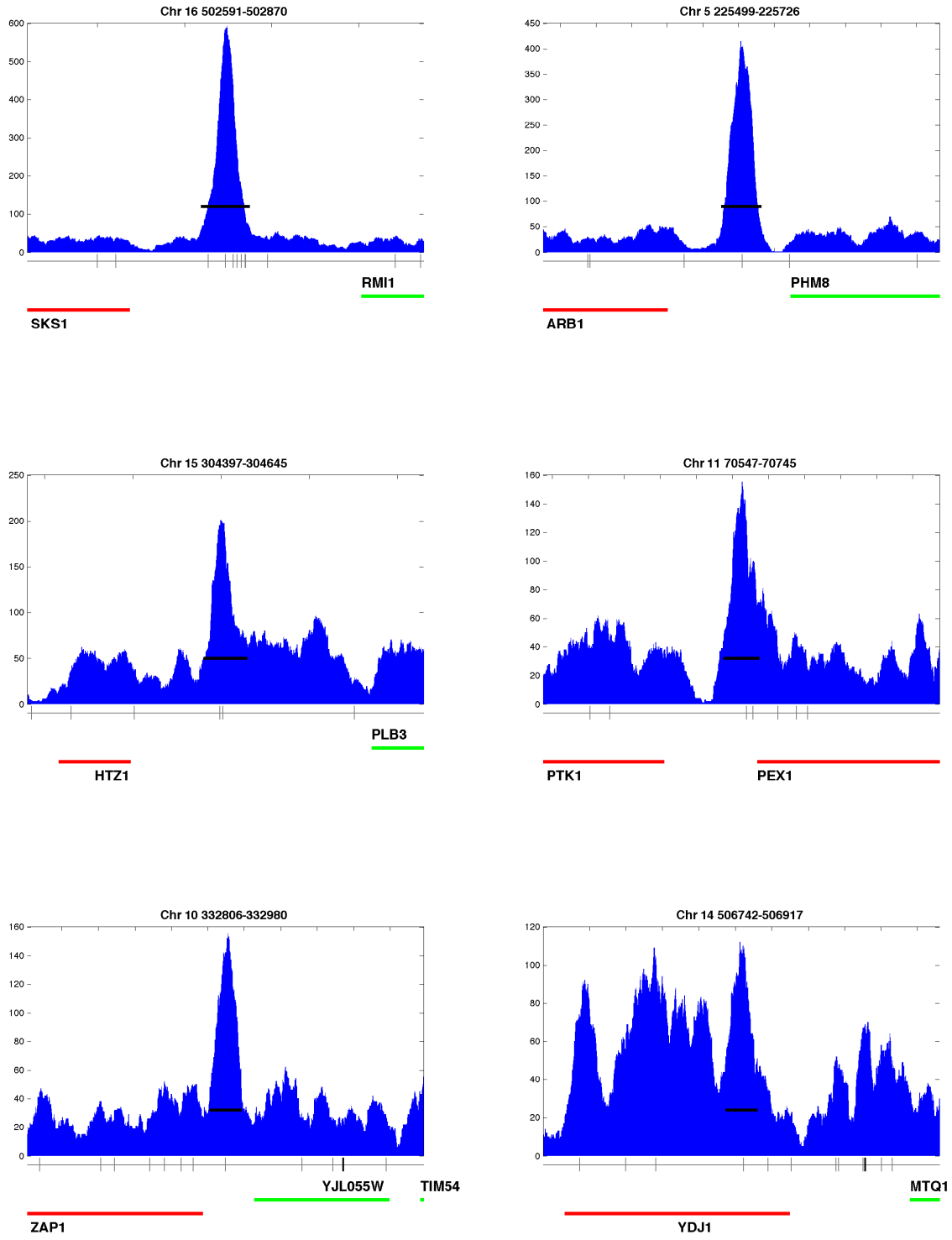
Supplemental Figure 1 (continued).



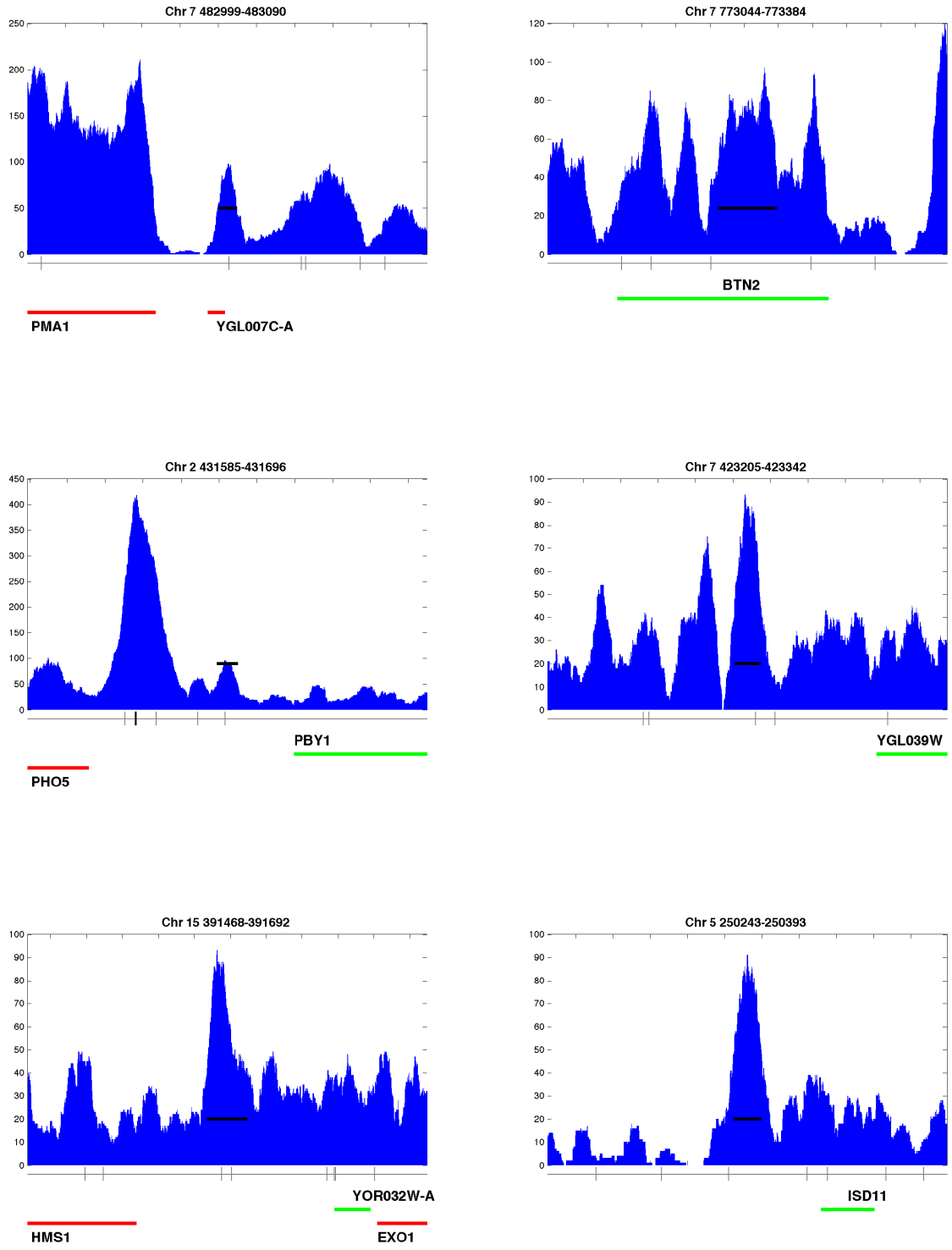
Supplemental Figure 1 (continued).

Appendix 2. Pho4-bound regions containing low affinity Pho4 binding sites.

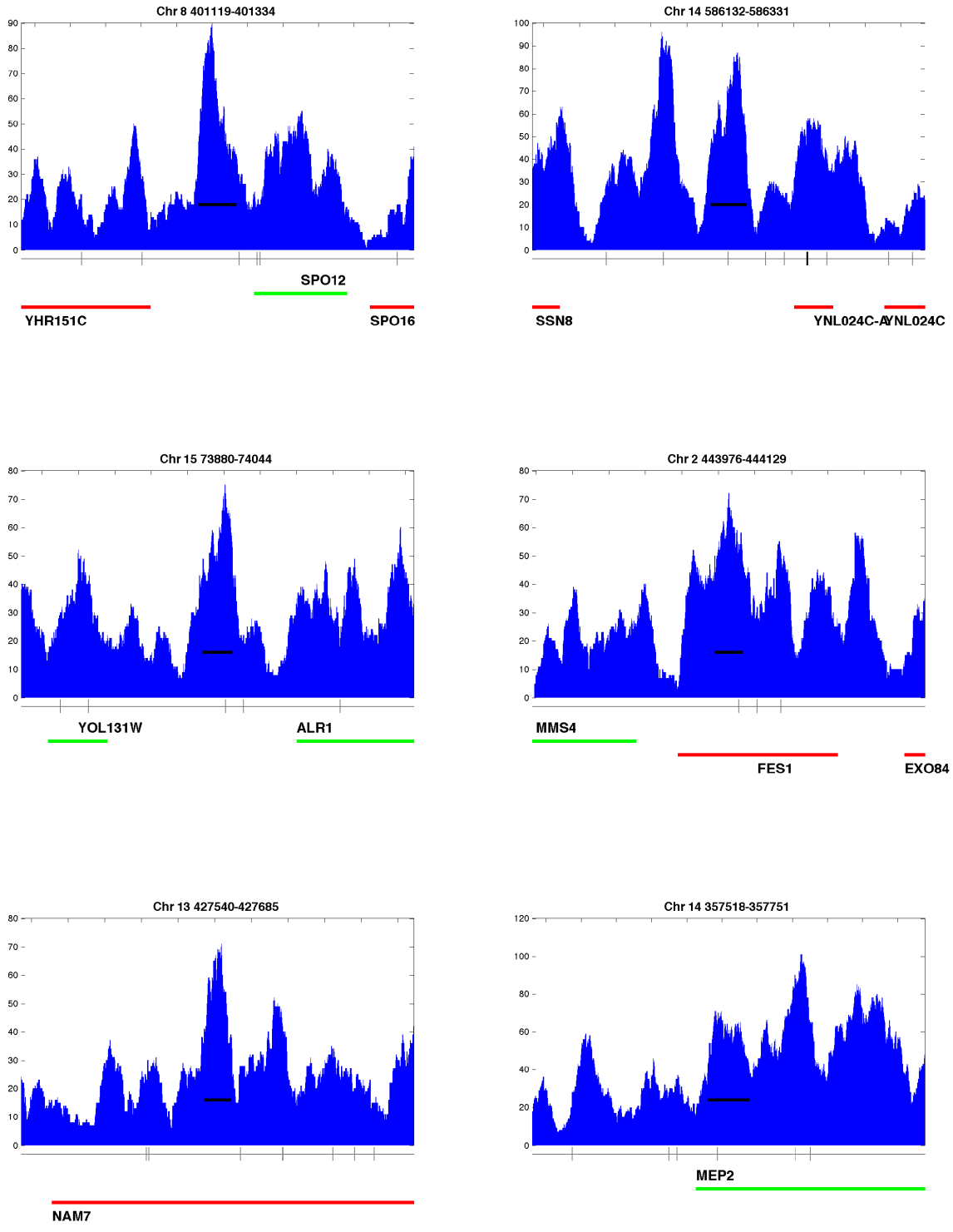
The plots show the Pho4 binding occupancy in no Pi conditions for the genomic region indicated above the plots. Black horizontal lines denote the regions that are determined as enriched of Pho4 binding ($p \leq 0.05$ for enrichment over genomic input and mock IP, enrichment $\geq$ 2-fold). The black vertical lines below the plots indicate the positions of high affinity binding sites, and the gray vertical lines indicate the positions of low affinity binding sites. The open reading frames on the forward (Watson) strand is colored in green and the ORFs on the reverse (Crick) strand is colored in red. The plots show 29 regions in total.



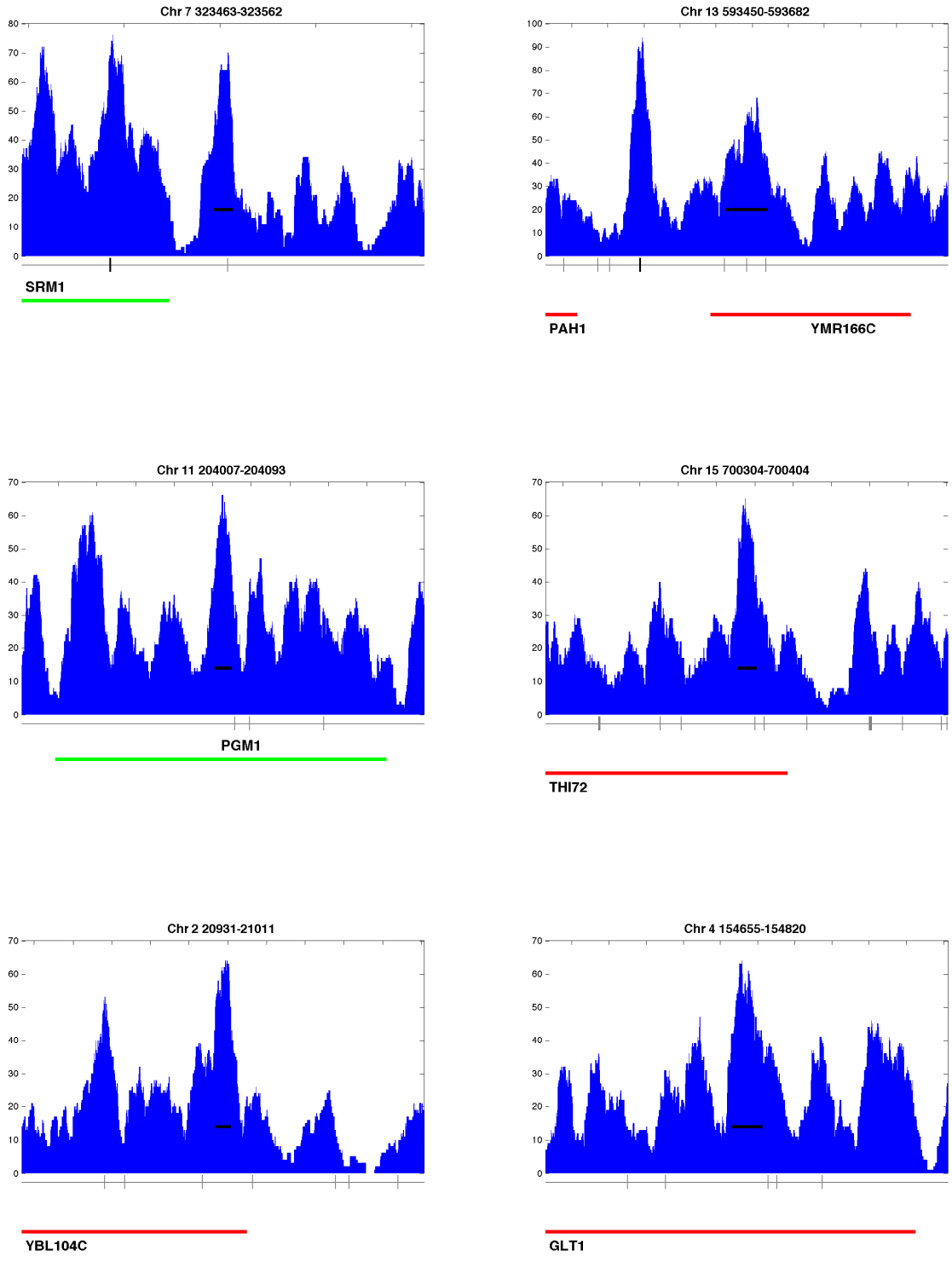
Supplementary Figure 2. Pho4-bound regions containing low affinity Pho4 binding sites



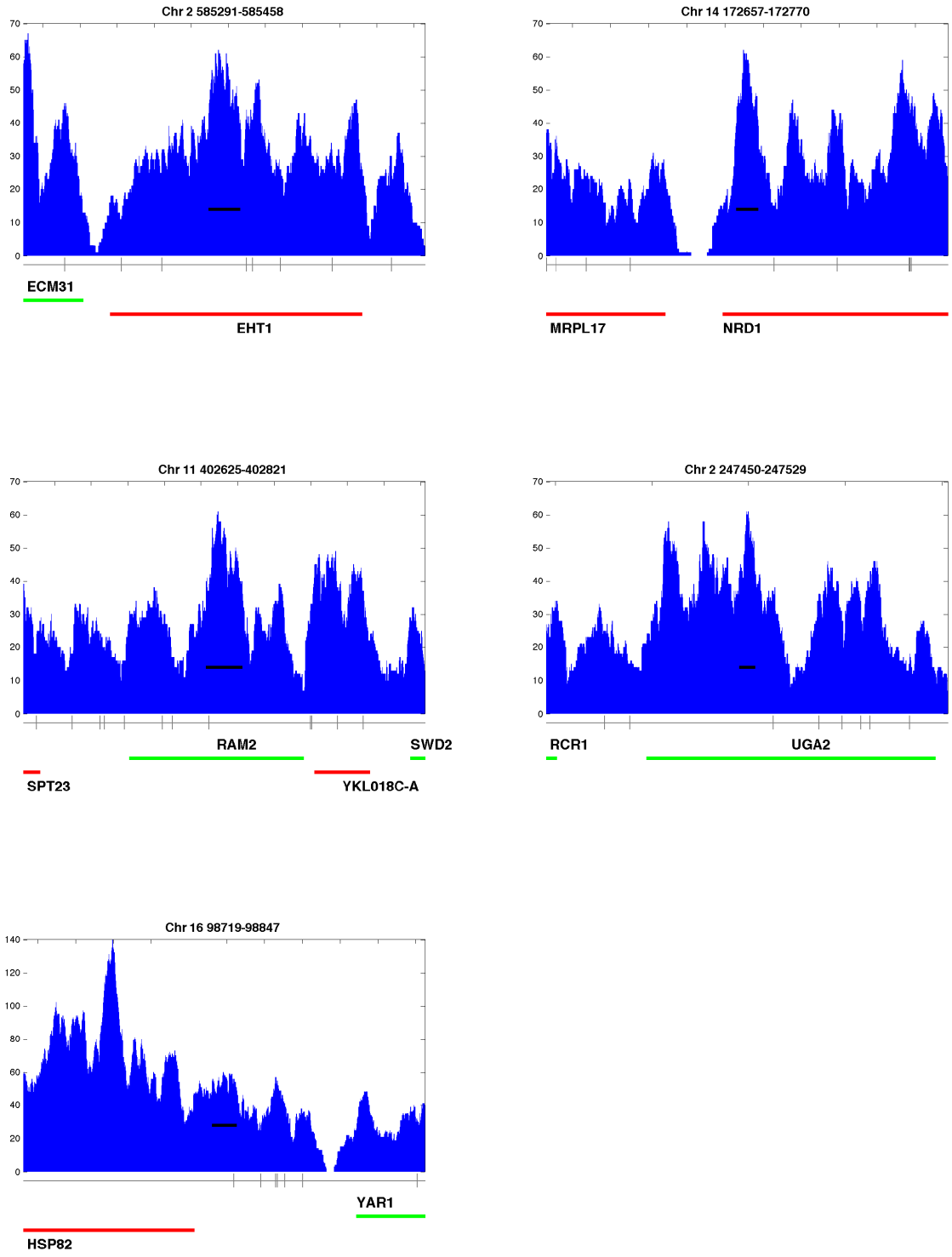
Supplemental Figure 2 (continued).



Supplemental Figure 2 (continued).



Supplemental Figure 2 (continued).



Supplemental Figure 2 (continued).

Appendix 3. List of all regions identified with Pho4 binding in no Pi conditions.

All the regions enriched with Pho4 binding are ranked based on the statistic significance (q-value and p-value) of their enrichment over mock-IP, and shaded with different colors based on thresholds of q-value $\leq 0.01, \leq 0.05$ and p–value $\leq 0.01, \leq 0.05$.

Chromosome	Start	End	Peak width	Peak height	p-value mock-IP	q-value mock-IP
13	25891	26454	564	6017	0.00E+00	0.00E+00
8	374925	375403	479	4253	0.00E+00	0.00E+00
16	517113	517525	413	3381	0.00E+00	0.00E+00
3	298939	299398	460	2609	0.00E+00	0.00E+00
10	413050	413366	317	1386	4.34E-133	1.64E-129
15	979011	979434	424	1177	1.75E-131	5.50E-128
6	131356	131647	292	1335	3.52E-124	9.48E-121
2	798651	799078	428	629	2.32E-122	5.48E-119
3	197148	197440	293	767	2.16E-90	4.54E-87
7	958428	958700	273	812	9.17E-80	1.73E-76
4	1022410	1022764	355	934	2.53E-70	4.33E-67
1	141735	141981	247	497	2.63E-54	4.14E-51
2	430996	431389	394	418	1.27E-52	1.84E-49
4	539138	539480	343	833	4.25E-52	5.73E-49
16	502591	502870	280	592	5.26E-40	6.62E-37
16	453527	453836	310	744	4.04E-39	4.76E-36
4	1420614	1420911	298	731	7.92E-37	8.79E-34
4	481731	482031	301	513	5.64E-33	5.92E-30
5	265873	266129	257	512	9.11E-32	9.04E-29
5	153028	153306	279	379	4.19E-26	3.95E-23
15	911899	912189	291	487	5.08E-25	4.56E-22
4	1345284	1345537	254	371	1.22E-24	1.04E-21
10	703349	703657	309	482	1.04E-22	8.54E-20
10	548158	548357	200	235	6.61E-21	5.19E-18
13	773740	773999	260	430	1.70E-19	1.28E-16
7	1000090	1000406	317	352	4.91E-19	3.57E-16
1	71131	71394	264	319	1.69E-18	1.18E-15
5	225499	225726	228	415	1.48E-17	9.99E-15
10	191831	192133	303	362	1.27E-16	8.29E-14
5	302549	302836	288	278	1.04E-15	6.56E-13
16	641219	641515	297	620	9.28E-15	5.64E-12
7	785643	786011	369	273	5.18E-14	3.06E-11
13	777528	777893	366	171	7.73E-13	4.29E-10
1	68974	69176	203	221	9.41E-13	5.07E-10
15	642499	642700	202	246	6.91E-12	3.62E-09
14	728237	728392	156	140	3.84E-10	1.90E-07
3	163514	163849	336	352	3.80E-10	1.94E-07

4	273835	274066	232	233	8.47E-10	4.10E-07
12	677610	677912	303	228	2.80E-09	1.29E-06
11	70547	70745	199	155	4.18E-09	1.83E-06
3	137104	137303	200	176	1.17E-08	5.02E-06
12	416911	417161	251	194	4.41E-08	1.81E-05
16	928347	928529	183	123	4.81E-08	1.93E-05
4	562890	563082	193	195	6.30E-08	2.48E-05
4	366704	366871	168	161	1.21E-07	4.65E-05
11	341761	341960	200	209	8.50E-07	3.21E-04
13	689199	689337	139	102	1.69E-06	6.27E-04
16	345151	345290	140	97	2.65E-06	9.61E-04
7	773044	773384	341	97	4.14E-06	1.45E-03
4	56784	56889	106	88	6.46E-06	2.21E-03
10	332806	332980	175	155	6.91E-06	2.29E-03
15	304397	304645	249	201	1.15E-05	3.62E-03
2	776295	776397	103	89	2.11E-05	6.23E-03
5	250243	250393	151	91	2.85E-05	8.28E-03
4	1084723	1084874	152	58	3.31E-05	9.32E-03
13	209129	209263	135	87	4.92E-05	1.35E-02
7	94668	94864	197	148	5.56E-05	1.48E-02
13	593450	593682	233	68	5.68E-05	1.49E-02
9	222702	222892	191	76	8.82E-05	2.19E-02
14	506742	506917	176	112	1.04E-04	2.48E-02
4	1454666	1454832	167	98	2.92E-04	4.92E-02
15	69175	69401	227	183	2.49E-04	4.95E-02
14	546066	546216	151	97	2.92E-04	4.97E-02
7	433330	433503	174	143	2.62E-04	5.00E-02
5	562718	562877	160	97	3.11E-04	5.09E-02
16	98719	98847	129	60	3.40E-04	5.48E-02
7	482999	483090	92	98	3.59E-04	5.73E-02
5	258208	258344	137	119	5.45E-04	7.45E-02
10	76680	76859	180	85	6.50E-04	8.45E-02
15	391468	391692	225	93	6.47E-04	8.47E-02
16	515982	516091	110	96	9.21E-04	1.08E-01
16	187881	188040	160	75	1.04E-03	1.15E-01
16	215390	215532	143	68	1.11E-03	1.16E-01
14	357518	357751	234	71	1.22E-03	1.22E-01
12	810496	810590	95	63	1.69E-03	1.44E-01
8	401119	401334	216	89	2.14E-03	1.64E-01

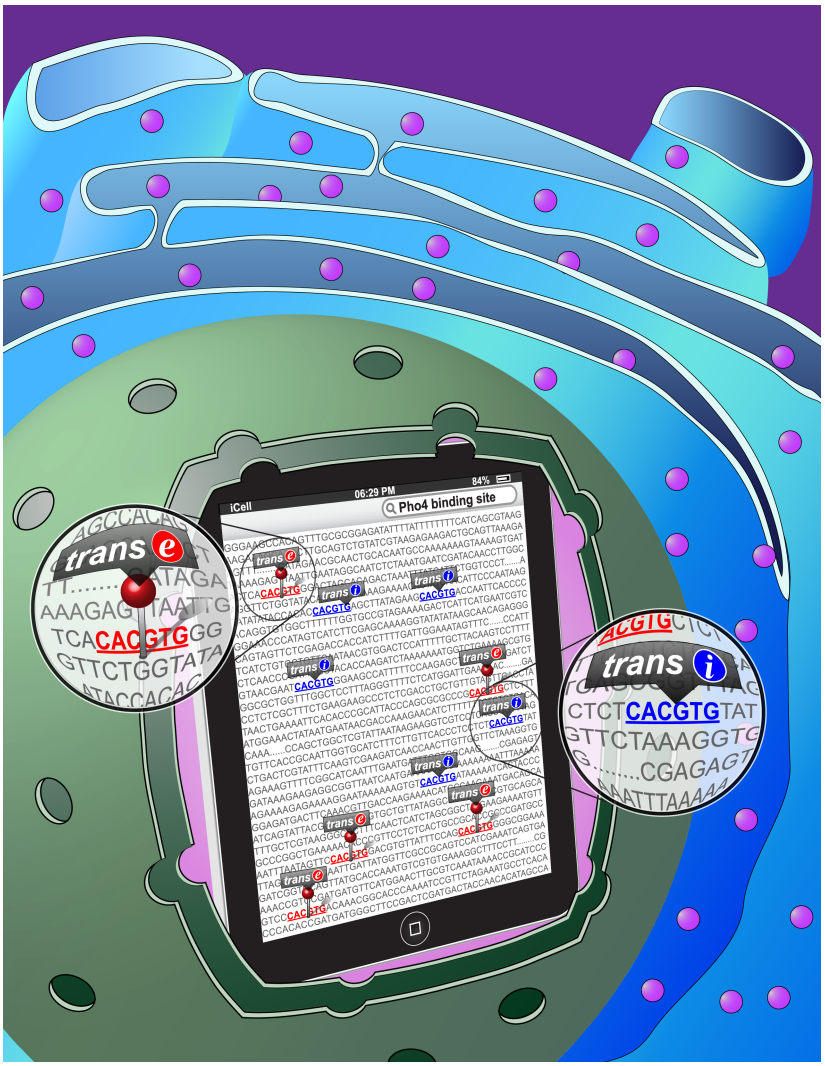
7	423205	423342	138	93	3.11E-03	1.90E-01
10	683144	683307	164	91	3.57E-03	1.97E-01
9	309757	309926	170	91	3.57E-03	1.98E-01
13	592920	593038	119	94	3.30E-03	1.98E-01
12	507031	507143	113	79	3.97E-03	2.09E-01
12	367572	367675	104	61	3.96E-03	2.10E-01
9	339755	339874	120	71	4.78E-03	2.17E-01
11	402625	402821	197	61	4.70E-03	2.18E-01
14	586132	586331	200	87	4.77E-03	2.18E-01
2	431585	431696	112	96	5.23E-03	2.37E-01
2	585291	585458	168	62	5.68E-03	2.38E-01
16	927387	927470	84	70	5.67E-03	2.38E-01
2	92152	92287	136	101	5.96E-03	2.46E-01
15	73880	74044	165	75	6.64E-03	2.60E-01
2	443976	444129	154	72	7.67E-03	2.74E-01
16	256334	256425	92	73	8.29E-03	2.81E-01
4	154655	154820	166	64	8.05E-03	2.85E-01
4	128779	128899	121	71	8.05E-03	2.85E-01
13	427540	427685	146	71	9.31E-03	3.03E-01
4	393447	393553	107	75	1.13E-02	3.31E-01
15	700304	700404	101	65	1.19E-02	3.32E-01
11	204007	204093	87	66	1.19E-02	3.35E-01
13	334986	335100	115	78	1.47E-02	3.48E-01
7	696727	696842	116	77	1.47E-02	3.49E-01
14	172657	172770	114	62	1.38E-02	3.49E-01
12	1030018	1030146	129	67	1.78E-02	3.70E-01
2	247450	247529	80	61	1.68E-02	3.73E-01
7	323463	323562	100	70	2.18E-02	3.88E-01
13	879196	879329	134	69	2.22E-02	3.92E-01
7	547119	547301	183	59	2.44E-02	4.21E-01
2	20931	21011	81	64	3.30E-02	4.53E-01
10	37304	37418	115	63	4.03E-02	4.81E-01

Appendix 4. Publication and art work

Integrated Approaches Reveal Determinants of Genome-wide Binding and Function of the Transcription Factor Pho4

Xu Zhou and Erin K. O'Shea

Molecular Cell. 42. 826 – 836. June 24, 2011



Supplementary Figure 3. Cover image for Molecular Cell 42. 826 – 836. June 24, 2011.

Designed by Xu Zhou and Erin O'Shea.

On the cover: Transcription factors recognize and bind to cognate binding sequences in the genome. However, consensus binding sites for transcription factors occur far more frequently than instances of genes bound or regulated by these factors. It is unclear what determines the sites to which a transcription factor binds and whether this binding is functional. Zhou and O'Shea describe how *trans* effects, such as competition with chromatin and competition and cooperativity with other transcription factors, shape the landscape of transcription factor binding and function. The cover shows an iPad search screen inside a eukaryotic nucleus (green) and surrounding endoplasmic reticulum (blue), on which consensus binding sequences for the transcription factor Pho4 (*cis* code, 'CACGTG') and *trans* effects are displayed embedded in genome sequence. *cis* and *trans* effects together determine whether consensus sites are bound (red font and stick-ball symbols) or not bound (blue font) by Pho4. The circled '*e*' and '*i*' represent different *trans* effects that either enhance (*e*) or inhibit (*i*) Pho4 binding.

Molecular Cell
Short Article

Integrated Approaches Reveal Determinants of Genome-wide Binding and Function of the Transcription Factor Pho4

Xu Zhou[1,2] and Erin K. O'Shea[1,2,3,*]
[1]Howard Hughes Medical Institute
[2]Department of Molecular and Cellular Biology
[3]Department of Chemistry and Chemical Biology
Harvard University Faculty of Arts and Sciences Center for Systems Biology, Northwest Labs, 52 Oxford Street, Cambridge, MA 02138, USA
*Correspondence: erin_oshea@harvard.edu
DOI 10.1016/j.molcel.2011.05.025

SUMMARY

DNA sequences with high affinity for transcription factors occur more frequently in the genome than instances of genes bound or regulated by these factors. It is not clear what factors determine the genome-wide pattern of binding or regulation for a given transcription factor. We used an integrated approach to study how *trans* influences shape the binding and regulatory landscape of Pho4, a budding yeast transcription factor activated in response to phosphate limitation. We find that nucleosomes significantly restrict Pho4 binding. At nucleosome-depleted sites, competition from another transcription factor, Cbf1, determines Pho4 occupancy, raising the threshold for transcriptional activation in phosphate replete conditions and preventing Pho4 activation of genes outside the phosphate regulon during phosphate starvation. Pho4 binding is not sufficient for transcriptional activation—a cooperative interaction between Pho2 and Pho4 specifies genes that are activated. Combining these experimental observations, we are able to globally predict Pho4 binding and its functionality.

INTRODUCTION

Specific transcriptional regulation is essential for precise control of many biological processes, including tissue development and responses to environmental stimuli. Alterations in gene expression are commonly associated with human diseases and disorders (Jimenez-Sanchez et al., 2001; Vaquerizas et al., 2009). A key regulatory step in the control of gene expression is the binding of transcription factors to DNA to activate gene transcription (Hochheimer and Tjian, 2003; Ptashne and Gann, 1997). Both *cis* and *trans* factors can influence transcription factor binding. In *cis*, binding typically requires recognition of a specific DNA sequence (Jacob and Monod, 1961); sequence variation in DNA binding motifs affects transcription factor binding occupancy (Kasowski et al., 2010; Schmidt et al., 2010; Zheng et al., 2010). However, high-affinity binding sites occur more frequently than experimentally detected binding events, even for transcription factors in organisms with a relatively small genome such as *Saccharomyces cerevisiae* (Harbison et al., 2004; MacIsaac et al., 2006). Among vertebrates, binding events display a species-specific pattern despite similarity between the consensus binding motifs of transcription factors (Schmidt et al., 2010). Among human individuals, the majority of transcription factor binding variance cannot be explained by genetic differences in binding sites (Kasowski et al., 2010), suggesting that *trans* factors also influence transcription factor binding. In *trans*, chromatin structure restricts DNA accessibility (Liu et al., 2006; Wasson and Hartemink, 2009), but the effect of nucleosome occupancy on the sequence-specific binding of transcription factors has not yet been systematically explored. Other *trans* factors, such as cooperating and competing factors (Pan et al., 2010; Pierce et al., 2003), have been studied primarily at the level of individual genes for their influence on transcriptional control, but not evaluated on a global scale. Moreover, binding of a transcription factor per se is often not sufficient for transcriptional regulation (Birney et al., 2007; Farnham, 2009; Harbison et al., 2004; MacIsaac et al., 2006). Thus, it remains unclear what determines the genomic locations to which a transcription factor binds and whether this binding is able to influence the transcription of a gene (Farnham, 2009).

To further complicate matters, transcription factors of the same family contain structurally conserved DNA binding domains and usually recognize similar short DNA motifs (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). However, these factors frequently regulate distinct biological responses (Robinson and Lopes, 2000; Sharrocks, 2001). The conflict between similarity in binding and divergence in regulation raises two questions: How are distinct patterns of transcriptional regulation achieved? How is the regulation by a given transcription factor influenced by others that recognize very similar DNA motifs? To answer these questions, it is necessary to systematically interrogate, on a genome-wide scale, the factors that may contribute to the specificity of transcription factor binding and regulation.

We used gene regulation by the phosphate-responsive (*PHO*) signaling pathway as a model system to explore the determinants for transcription factor binding and function. Pho4, a basic-helix-loop-helix (bHLH) transcriptional activator in

Supplementary Figure 4. Print of publication.

S. cerevisiae, is regulated in response to environmental inorganic phosphate (Pi) availability and activates a transcriptional program together with the transcription factor Pho2 when cells are limited for Pi (Ogawa et al., 2000). In Pi rich conditions, Pho4 is phosphorylated and transported from the nucleus into the cytoplasm; in phosphate-limited conditions, Pho4 is dephosphorylated and transported from the cytoplasm into the nucleus (O'Neill et al., 1996; Schneider et al., 1994). Pho4 binds with high affinity to "CACGTG" motifs in vivo (Harbison et al., 2004) and in vitro (Badis et al., 2008; Maerkl and Quake, 2007; Zhu et al., 2009). In this study, we found that Pho4 is bound to only ~14% of its consensus binding sites, and only a quarter of those binding events result in activation of gene transcription. We demonstrated that this genome-wide binding and regulatory pattern is determined by a combination of chromatin restriction, competitive binding from the factor Cbf1 that recognizes the same motif as Pho4, and cooperation with Pho2. We also discovered that Cbf1 competition raises the threshold of gene activation and helps to define the specificity of *PHO* gene regulation. Combining our experimental observations, we can predict transcription factor binding and function at a whole genome level.

RESULTS

Chromatin Negatively Regulates Pho4 Binding

We applied biotin-tagging chromatin immunoprecipitation (Kolodziej et al., 2009; van Werven and Timmers, 2006) combined with high-throughput sequencing (Bio-ChIP-Seq; see the Experimental Procedures) to identify Pho4 binding events in vivo in phosphate starvation conditions (Figure 1A and Figure S1A available online). Although there are 843 "CACGTG" consensus binding sites (all "NCACGTGN" motifs except "TCACGTGA" are considered to be consensus binding sites; see the Experimental Procedures for more details) in the genome, we observed Pho4 binding to only 115 (~14%) of these sites (Experimental Procedures), implying that factors other than DNA binding specificity influence Pho4 binding in vivo. Since chromatin restricts the access of transcription factors to their potential binding sites (Khorasanizadeh, 2004; Kornberg and Lorch, 1999; Narlikar et al., 2002), a Pho4 "CACGTG" consensus binding site could exist in one of two possible states: an inaccessible state, occluded by nucleosomes; or an accessible state, exposed in a nucleosome-depleted region (nucleosome-free or nucleosome linker region). To determine whether local chromatin structure influences transcription factor binding, we mapped nucleosome occupancy in no-Pi conditions with micrococcal nuclease digestion followed by paired-end deep sequencing (Figure 1B, Figures S1B–S1E, and the Experimental Procedures). As expected, Pho4 is not bound to the binding sites that are inaccessible and most occluded by nucleosomes (see the Experimental Procedures for details) (209 of 216 sites, 97%; Figure 2A and Figure S2A). However, with the same threshold, Pho4 is also not bound to two thirds of the most accessible sites (172 of 248 sites) (Figure 2A and Figure S2A). We conclude that chromatin structure inhibits transcription factor binding, but open chromatin structure is not sufficient for Pho4 binding at its "CACGTG" consensus sites.

Competition from Cbf1 Determines Pho4 Occupancy In Vivo at Nucleosome-Depleted Sites

It is possible that proteins with similar specificity compete with Pho4 for binding to accessible "CACGTG" consensus sites. Cbf1, another member of the bHLH transcription factor family, is present in the nucleus at high concentration (Ghaemmaghami et al., 2003), is not known to interact with Pho4 (Graumann et al., 2004), and binds with high affinity to the same consensus binding motif "CACGTG" in vitro and in vivo as does Pho4 (Harbison et al., 2004; MacIsaac et al., 2006; Maerkl and Quake, 2007; Zhu et al., 2009). To test whether Cbf1 competes with Pho4 for binding, we identified in vivo binding sites for Cbf1 in high- and no-Pi conditions. In no-Pi conditions, 77% (132 of 172) of the accessible consensus "CACGTG" binding sites not bound by Pho4 are occupied by Cbf1 (Figure 2B). Intriguingly, most of the accessible sites bound by Pho4 (72 of 76, 95%) are also bound by Cbf1. Thus, the accessible, high-affinity "CACGTG" sites mainly fall into two classes: those in which Cbf1 competes with Pho4 most effectively, resulting in detectable binding of Cbf1 but not Pho4, and those in which Cbf1 competes less effectively, resulting in significant occupancy of both Pho4 and Cbf1. In high-Pi conditions, Cbf1 is bound to both of these classes of sites (119 of 132 and 67 of 72, respectively; Figure S2B).

Pho4 and Cbf1 have different preferences for bases flanking the "CACGTG" consensus binding site in vitro (Maerkl and Quake, 2007); these sequence features might explain differences in Cbf1 and Pho4 occupancy in vivo. In accord with the observed in vitro sequence preferences, we find that accessible sites with less Pho4 binding have a single 5′ "T" base flanking the "CACGTG" (Figure 2C), suggesting that Cbf1 can compete most effectively at these sites (Figure 2D).

Competitive Binding of Cbf1 Influences the Activation Threshold and Specificity of the *PHO* Regulon

What is the physiological role of the interplay between Pho4 and Cbf1 at consensus sites? One idea is that Cbf1 is required in high-Pi conditions to keep nucleosomes properly positioned, and therefore keep consensus binding sites nucleosome free (Figure S3A; "Cbf1 priming model"). Although there is precedent for Cbf1 positioning nucleosomes in the promoters of some sulfur metabolism genes (Kent et al., 2004; Kent et al., 1994), when we analyzed the *cbf1Δ* strain in high-Pi conditions we observed no change in nucleosome occupancy at Pho4 binding sites in Pho4-regulated genes ($r = 0.943$ between *cbf1Δ* strain and wild-type, $r = 0.953$ between wild-type replicates; Figure S3B); in contrast, in the *cbf1Δ* strain nucleosome occupancy increases and nucleosome position shifts at binding sites in the regulatory regions of sulfur metabolism genes (Lee et al., 2010) (Figure S3C).

An alternative model is that competition from Cbf1 prevents spurious activation of phosphate-responsive genes in high-Pi conditions, and of other "CACGTG"-containing genes during phosphate starvation (Figure S3D; "Cbf1 blocking model"). In the absence of Cbf1, most Pho4-regulated genes (see definition in Figure 4A) showed significantly increased expression in high-Pi medium (20 of 28, $p \leq 0.05$; Figure 3A, column 2)—conditions in which Pho4 is less active and localized primarily to the cytoplasm (Komeili and O'Shea, 1999). This aberrant expression is Pho4 dependent (compare Figure 3A, columns 2 and 3) and is

Supplementary Figure 4 (continued).

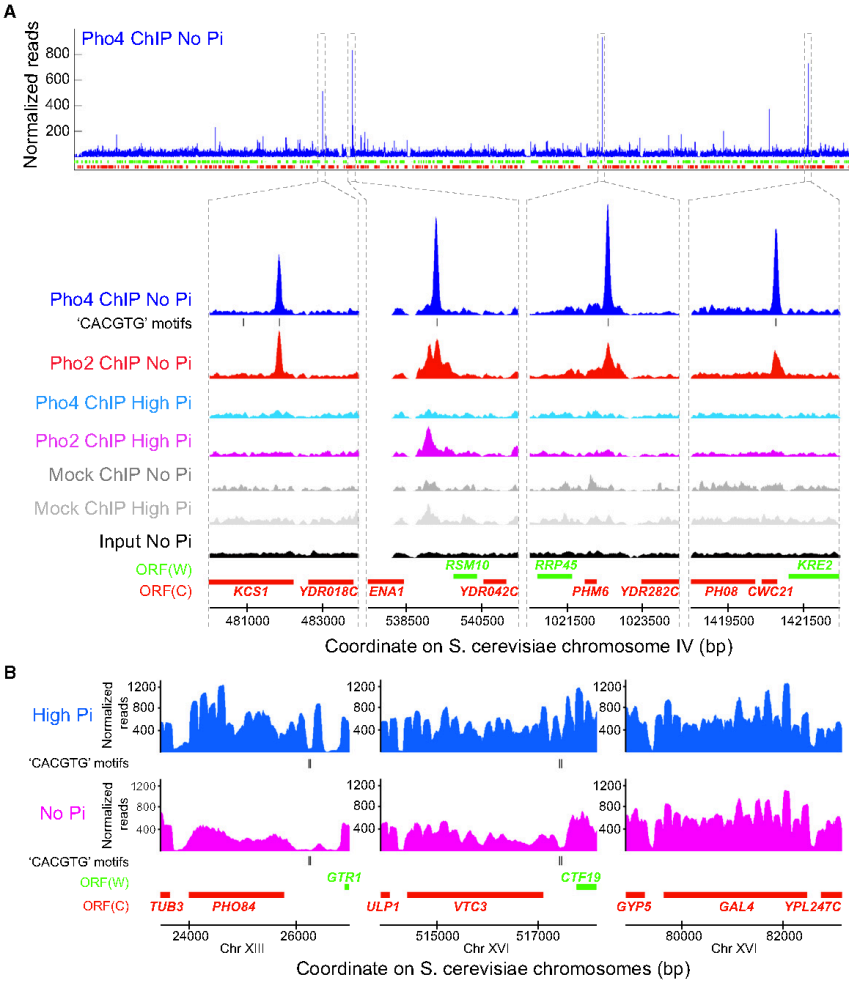


Figure 1. ChIP-Seq Analysis and Examples of Nucleosome Positioning in High- and No-Pi Conditions

(A) ChIP-seq analysis of Pho2 and Pho4 binding and a mock IP (wild-type strain with integrated *E.coli birA* gene) in high- and no-inorganic-phosphate (Pi) conditions, and genomic input in no-Pi conditions. All ChIP results were normalized to Pho4 ChIP at no-Pi conditions (Supplemental Experimental Procedures).

(B) Nucleosome occupancy map for *PHO84* and *VTC3* (both derepressed in no-Pi conditions), *GAL4* (not responsive to Pi concentration), in high- and no-Pi conditions. The nucleosome maps in both conditions were normalized to have the same number of total reads.

See also Figure S1.

not the indirect result of an upstream signaling defect caused by deletion of Cbf1 (Figure S3E). Moreover, deletion of Rtg3 and Tye7, two other members of the bHLH family that bind the "CACGTG" consensus site (Zhu et al., 2009), did not result in spurious activation (Figure 3A, column 4). Since deletion of Cbf1 causes a growth defect and impaired activation of the *PHO* pathway in no-Pi conditions (Figure S3E), we evaluated the consequences of Cbf1 competition when the *PHO* pathway is fully activated using a strain lacking the cyclin Pho80 grown in high-Pi medium (O'Neill et al., 1996). In the *cbf1⊿ pho80⊿* strain, we observed Pho4 binding to sites bound by Cbf1 in the wild-type strain (Figure 3D), particularly at the "T-CACGTG"

Supplementary Figure 4 (continued).

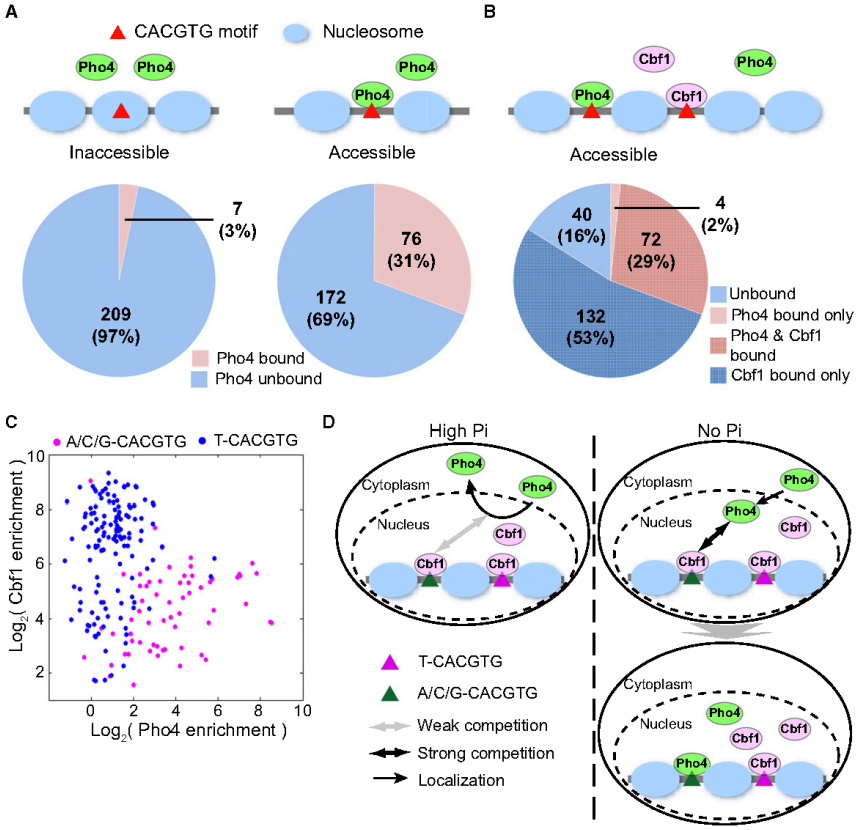


Figure 2. Determinants of Pho4 Binding at Its Consensus Binding Motifs

(A) Pie charts showing Pho4 binding at inaccessible (the quartile most occluded by nucleosomes) and accessible (the quartile least occupied by nucleosomes) consensus binding sites in no-Pi conditions.

(B) Pie chart showing Pho4 and Cbf1 binding at accessible consensus binding sites in no-Pi conditions.

(C) Scatter plot displaying Pho4 and Cbf1 binding enrichment in no-Pi conditions for accessible sites that are bound by at least one of the two transcription factors.

(D) Schematic depicting a model for determinants of Pho4 binding to its consensus motifs. In high-Pi conditions (left), Pho4 is imported into the nucleus and then actively exported; thus, the Pho4 nuclear concentration is low (Komeili and O'Shea, 1999; O'Neill et al., 1996). Most of the accessible consensus binding sites are occupied by Cbf1, which resides in the nucleus constitutively (Huh et al., 2003). In no-Pi conditions (upper right), Pho4 is no longer exported from the nucleus and the nuclear concentration of Pho4 increases, allowing Pho4 to compete effectively for binding at sites that are weakly bound by Cbf1 (lower right); Cbf1 is bound to the consensus sites with a 5′ flanking "T" with high enough affinity to prevent Pho4 binding.

See also Figure S2 and Table S2.

motifs (Figure 3C and Figure S3F). Indeed, Pho4 is bound to and activates transcription of 13 genes that were bound by Cbf1 and not regulated by Pho4 in the wild-type strain (Figure 3B). Our observations support two roles for Pho4-Cbf1 competition: in high-Pi conditions, Cbf1 prevents spurious activation of the *PHO* genes induced by a low level of nuclear Pho4, ensuring that phosphate-responsive genes are turned off when Pi is available; in no-Pi conditions, Cbf1 prevents Pho4 from inappropriately activating genes containing a "CACGTG" motif that are not part of the phosphate regulon, ensuring that only genes needed for the response to phosphate limitation are turned on.

Cooperative Binding between Pho2 and Pho4 Determines the Functionality of Pho4 Binding Events

Pho4 binds to the "CACGTG" motif in the promoters of over 80 genes, but only ~10–20 genes are regulated by Pho4 in

Supplementary Figure 4 (continued).

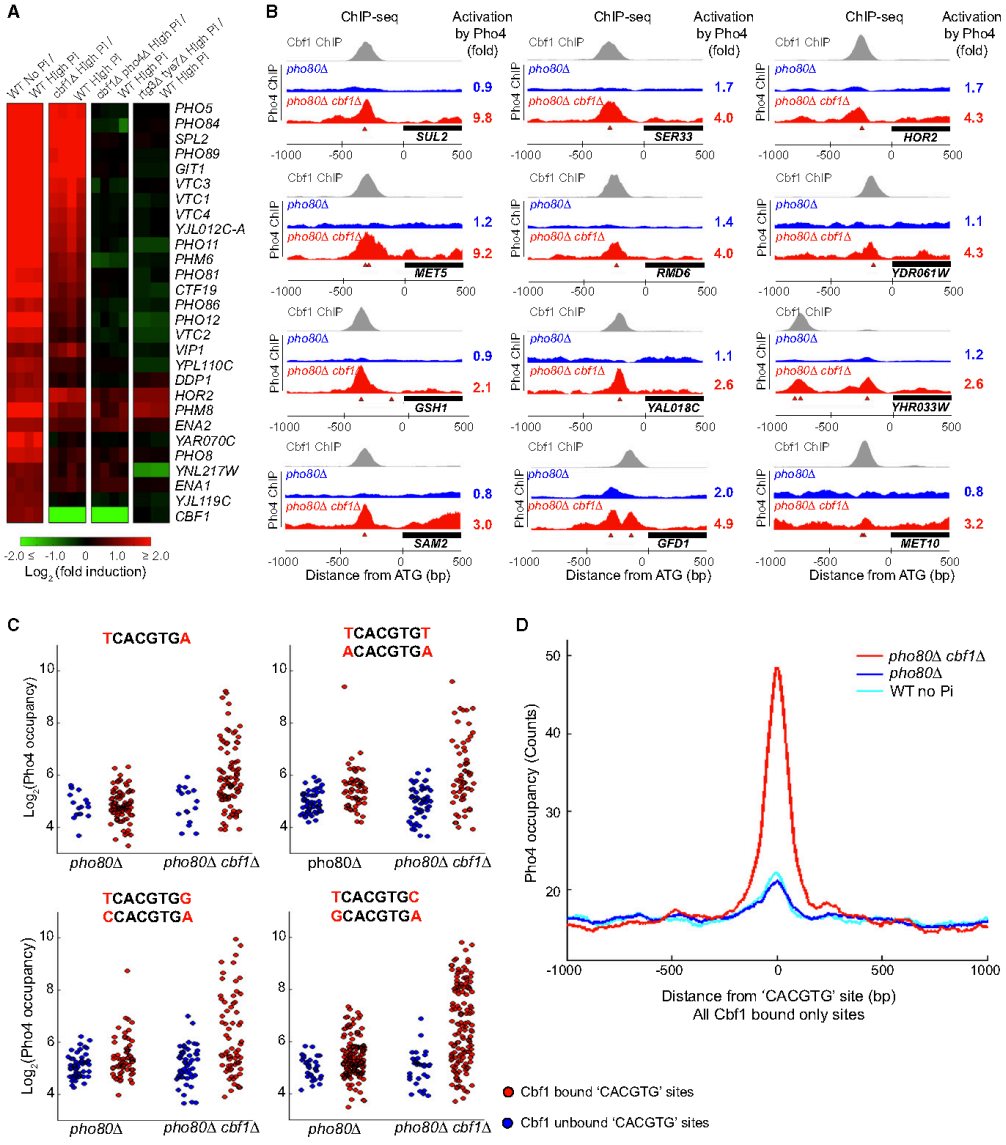


Figure 3. Cbf1 Prevents Pho4 Spurious Transcriptional Sctivation in High-Pi Conditions and Inappropriate Gene Activation in No-Pi Conditions

(A) Heat map showing the fold induction (log_2 scale) of Pho4-regulated genes for wild-type in no-Pi conditions (column 1) and *cbf1Δ*, *cbf1Δ pho4Δ*, and *rtg3Δ tye7Δ* strains in high-Pi conditions (columns 2–4), as measured by microarray analysis. Subcolumns indicate biological replicates.

(B) Differential binding and activation by Pho4 in the presence and absence of Cbf1. Tracks show ChIP-seq results of Cbf1 binding in no-Pi conditions (gray), Pho4 binding in a *pho80Δ* strain (constitutively nuclear Pho4, blue) and in a *cbf1Δ pho80Δ* (red) strain. Gene activation by Pho4 in *pho80Δ* (blue) and *cbf1Δ pho80Δ*

Supplementary Figure 4 (continued).

response to Pi starvation (Ogawa et al., 2000; Springer et al., 2003). Moreover, these regulated genes have different dependence on Pho4 and on Pho2, a homeodomain factor that interacts with Pho4 and regulates the phosphate starvation response (Vogel et al., 1989). We used epistasis expression analysis (mutant cycle analysis) (Capaldi et al., 2008) to dissect and quantify the contribution of the regulatory interactions between Pho2 and Pho4 to transcriptional activation (Figure S4) in terms of three "expression components": the contribution of Pho2 acting alone (Pho2), the contribution of Pho4 acting alone (Pho4), and the contribution of Pho2 and Pho4 acting together (which we refer to as a "cooperative" component). To determine the values of these expression components, we directly compared gene expression between all possible pairs of wild-type, *pho2Δ*, *pho4Δ*, and *pho2Δ pho4Δ* strains in no- and high-Pi conditions, in which the measured gene expression difference between each pair of strains consists of a linear sum of the expression components relevant for that pairwise comparison (Figure S4). We calculate the expression components for each gene through regression methods with equations describing the expression components measured in each microarray (Supplemental Experimental Procedures). We observe only three regulatory interactions operating at phosphate-responsive genes: gene activation by Pho4 acting alone (Pho4 only), genes activated by the interaction between Pho2 and Pho4 (Cooperative, Co), and mixed regulation (genes partially activated by Pho4 alone and partially dependent on the interaction between Pho2 and Pho4, Pho4 + Co) (Figure 4A). Notably, almost all Pho4 regulated genes contain a cooperative component, suggesting that the interaction between Pho2 and Pho4 is necessary for gene activation.

Since Pho2 and Pho4 interact and bind cooperatively to the *PHO5* promoter (Barbaric et al., 1998; Barbarić et al., 1996; Vogel et al., 1989), we hypothesized that this cooperative binding might be correlated with the functionality of Pho4 binding events—the ability to trigger gene activation. We observed a strong correlation between Pho2 and Pho4 binding events in no-Pi conditions ($r = 0.926$, $p < 10^{-49}$; Figure S5A). Only some of these coincident binding events appear to be instances of cooperative binding (see the Experimental Procedures), where both Pho2 and Pho4 occupancy increases in response to phosphate starvation—it is these cooperative binding events that correlate with gene activation ("regulated" class in Figure 4B, $p = 8.1 \times 10^{-16}$, Fisher's exact test). Recruitment of Pho4 to cooperative, regulated sites is entirely dependent on Pho2, whereas Pho4 binding to the noncooperative, nonregulated sites is largely unaffected by deletion of Pho2 (Figure 4C). The reduction of Pho4 occupancy in *pho2Δ* strains in no-Pi conditions is correlated with the transcriptional activation contributed by the cooperative interaction between Pho2 and Pho4 (Co component) for Pho4-regulated genes ($R = 0.63$, $p = 0.0016$). For the cooperative Pho4 binding events, the Pho2 and Pho4 ChIP signals are overlapping (Figure 4D), and the predicted Pho2 binding sites are enriched at a distance of 15 bp from the consensus Pho4 binding sites (Figure 4E). In contrast, when only one of the factors is significantly recruited in response to Pi limitation, the gene is not activated ("nonregulated" class in Figure 4B, $p = 0.99$, binomial test), there is no juxtaposition of Pho2 and Pho4 ChIP signals (Figure 4D), and predicted Pho2 binding sites do not exhibit a consistent spatial relationship with Pho4 binding sites (Figure 4E). Thus, the spatial organization of Pho2 and Pho4 binding motifs may promote cooperative binding of Pho2 and Pho4 and the ability to activate transcription.

Prediction of Pho4 Binding and Function

Prediction of transcription factor binding and functional targets has been a challenging task. Here, we provide an integrated mechanistic view of the determinants that influence Pho4 binding and regulation, taking into account all "CACGTG" sites independent of evolutionary conservation, clustering of motifs, and relative positioning in the promoter. Incorporating the influence of *trans* effects into an equilibrium binding model (see the Experimental Procedures), 43 of 50 (86%) (Figure 4F) of the top predicted binding sites are indeed bound by Pho4 (AUC-ROC = 0.87). We find that all *trans* effects contribute significantly to the pattern of Pho4 binding in the genome (Figure S5C). Either nucleosome occupancy or flanking sequences predicts Pho4 binding to high-affinity consensus sites (top panel; AUC-ROC). However, if we consider only the group of top predicted targets (bottom two panels), a prediction based on nucleosome occupancy or Cbf1 competition alone has poor accuracy, whereas the synergy between Cbf1 competition and nucleosome occlusion more accurately predicts binding to this group of target sites (Figure S5C and the Supplemental Experimental Procedures). Of the 115 experimentally determined Pho4 binding sites, fewer than 25% are able to promote activation of gene transcription in no-Pi conditions. Considering the cooperative interaction between Pho2 and Pho4, 23 of 28 (82%) (Figure 4F) binding sites predicted to be functional are actually associated with transcriptional activation (AUC-ROC = 0.992). From this prediction, we identified three nonpromoter binding events that are associated with Pi-dependent antisense transcripts (B. Zid and E.K.O., unpublished data)—two of them have not been previously identified and are potentially linked to phosphate signaling. We find that the presence of a Pho2 binding motif 15 bp from the Pho4 binding site is highly predictive of the functionality of the Pho4 binding events (Figure S5D), indicating that the spatial arrangement of Pho2 and Pho4 binding sites is critical in activating gene transcription.

(red) strains is determined by comparing gene expression in *pho80Δ* and *pho80Δ pho4Δ* strains and *cbf1Δ pho80Δ* and *cbf1Δ pho80Δ pho4Δ* strains, respectively. Red triangles mark the consensus "CACGTG" binding sites.
(C) Scatter plots show Pho4 binding occupancy at all "TCACGTG" sites, both accessible and those occluded by nucleosomes. Labeled 8-mer sequences indicate DNA motifs with the same binding preference of Pho4.
(D) Average Pho4 binding occupancy at sites that are Cbf1-bound only in the wild-type.
See also Figure S3 and Table S2.

Supplementary Figure 4 (continued).

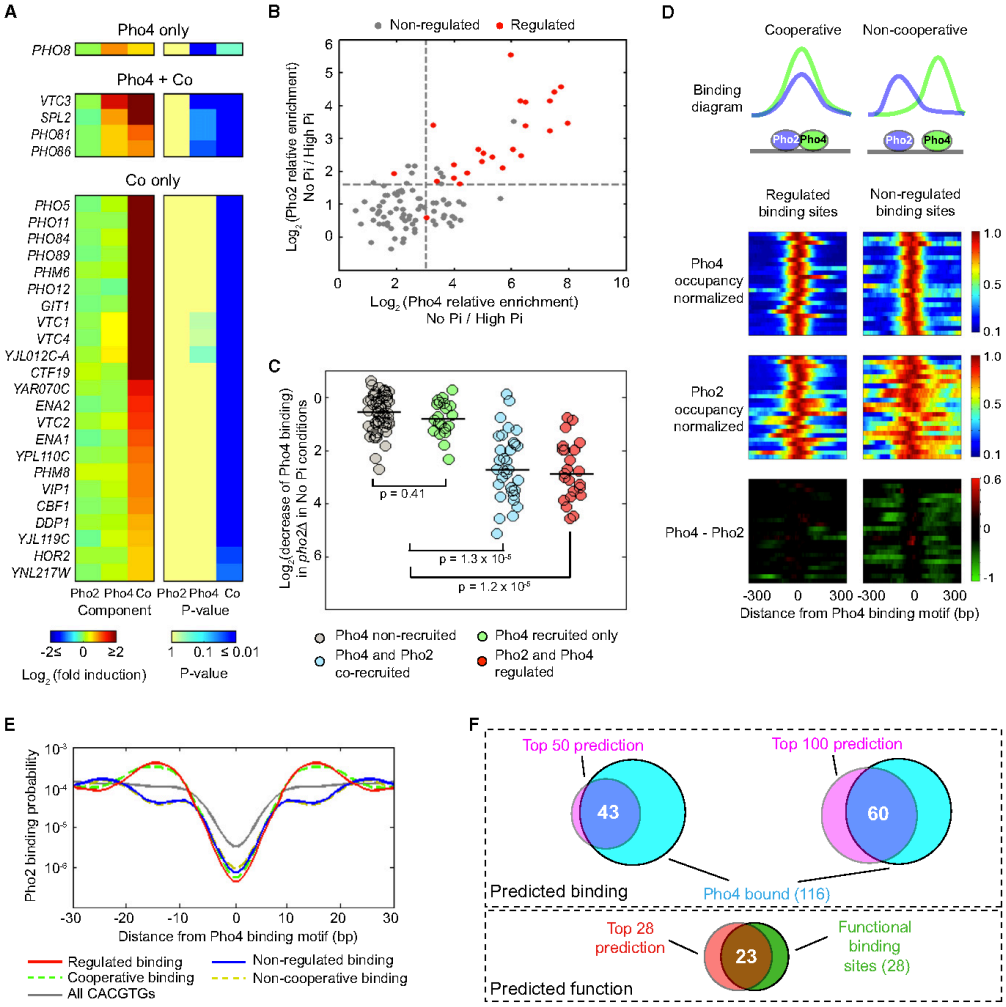


Figure 4. Mutant Cycle Analysis and Correlation between Pho2 and Pho4 Binding and Gene Induction

(A) Heat map showing the best fit of expression components (left columns) and their statistical significance (right columns) for genes that regulated by Pho2 and Pho4. Genes are clustered according to the statistical significance ($p \leq 0.05$) of the Pho2, Pho4, and cooperative component (Co).

(B) Scatter plot showing recruitment ($enrichment_{high}/enrichment_{no}$) of Pho2 and Pho4 after Pi starvation for all Pho4-bound consensus sites within 800 bp of the transcription start site (TSS). "Regulated binding sites" are sites associated with Pho4 regulated genes (Figure 4A). Gray dashed lines indicate the recruitment threshold as mean + 2 standard deviations (SDs) for all consensus motifs excluding the regulated sites.

(C) Plots showing fold decrease in Pho4 occupancy at Pho4-bound consensus binding sites in *pho2Δ* strains. Black lines indicate the median, and p values are calculated with two-sample t tests with unequal variance.

(D) Heat map displaying normalized ChIP occupancy of Pho2 and Pho4 in no-Pi conditions for all sites showing Pho4 recruitment (≥ mean + 2 SDs).

(E) Spatial organization of Pho2 binding motifs predicted based on in vitro binding specificity (Zhu et al., 2009).

(F) Venn diagram showing prediction of Pho4 binding and function derived from a model that incorporates competition and cooperativity.

See also Figures S4 and S5 and Tables S2 and S3.

Supplementary Figure 4 (continued).

DISCUSSION

Sequence-specific transcriptional factors recognize short *cis* elements in promoter regions to regulate gene transcription (Farnham, 2009; Jacob and Monod, 1961). However, these elements exist tens of times more frequently than the number of in vivo binding events. How can transcription factors bind to and regulate a specific set of genes when many other genes carry the same binding sites? We used a genome-wide approach to systematically dissect the determinants of transcription factor binding and regulation, and investigated whether these determinants can explain transcriptional specificity. We provide evidence that competition from chromatin, competition from another transcription factor that recognizes Pho4 consensus binding motifs, and cooperativity with a partner protein determine the specificity of Pho4 binding and regulation. We also discovered novel regulatory functions for Cbf1 competition and provide a general view of the specific regulation of the *PHO* regulon.

In high-Pi conditions, Pho4 is phosphorylated and exported to the cytoplasm (O'Neill et al., 1996; Schneider et al., 1994). Competition from Cbf1 increases the threshold of transcriptional activation by Pho4, preventing spurious activation by low levels of nuclear Pho4 (Figure 2D). In no-Pi conditions, phosphorylation of Pho4 is inhibited (O'Neill et al., 1996; Schneider et al., 1994) and the increase in the concentration of unphosphorylated nuclear Pho4 enables it to compete effectively with Cbf1 at sites where Cbf1 is weakly bound—those with certain flanking sequences found in genes within the phosphate regulon; inappropriate binding and activation of "CACGTG"-containing genes outside the phosphate regulon is prevented. The dynamic range of binding and gene activation is increased as a result of the reduction in Pho4 binding due to Cbf1 competition in high-Pi conditions and its condition-dependent cooperative interaction with Pho2 in no-Pi conditions (Pho2 interacts only with unphosphorylated Pho4 during phosphate limitation) (Komeili and O'Shea, 1999). Other stress- and nutrient-response transcriptional programs may employ similar strategies to trigger high levels of induction of specific genes that were tightly repressed in the absence of the perturbation.

In mammalian cells, transcription factors within the same structural family commonly recognize similar DNA motifs but regulate diverse biological processes (Badis et al., 2009; Berger et al., 2008; Wei et al., 2010). It is unclear how faithful and distinct regulation is achieved with transcription factors of overlapping specificity. Our results suggest that chromatin structure influences the accessibility of potential binding sites and competition among transcription factors can facilitate discrimination of subtle differences in DNA binding specificity across the genome. Cooperation with conditionally, spatially or temporally expressed coactivators (Sharrocks, 2001) may provide additional selection for binding events to elicit specific transcriptional outcomes. Moreover, competition among transcription factors may have significant effects on transcriptional regulation, controlling the threshold for gene activation and limiting potential crosstalk between different signaling pathways. Overall, our findings suggest that the specificity of transcriptional regulation is a composite of the DNA binding specificity, nuclear abundance and functional interactions of each transcription factor in the context of eukaryotic genome architecture. Understanding the grammar of specific regulation of individual transcription factors will provide insights into decoding the complex regulatory network of the eukaryotic genomes.

EXPERIMENTAL PROCEDURES

Strains

Methods of strain construction are described in the Supplemental Experimental Procedures. All strains used in this study are listed in Table S1.

Media and Growth Conditions

Media and growth conditions are described in the Supplemental Experimental Procedures.

Defining the Consensus Pho4 Binding Motif

High-affinity Pho4 binding motifs were determined with the position specific scoring matrix (PSSM) described in Lam et al. (2008), which was derived from in vitro measurement of Pho4 DNA binding affinities (Maerkl and Quake, 2007). We selected the most stringent threshold (0.0075) to recapitulate in vivo validated high-affinity binding sites as the threshold of Pho4 high-affinity binding motifs (Lam et al., 2008). Since all determined high-affinity binding motifs at this threshold contain "CACGTG" as core sequence, we define these motifs as the consensus "CACGTG" binding motif for Pho4. All "NCACGTGN" motifs except "TCACGTGA" in *S. cerevisiae* genome meet the threshold.

Biotin-Tagging Immunoprecipitation with High-Throughput Sequencing

The procedure for biotin-tagging immunoprecipitation is described in detail in the Supplemental Experimental Procedures. In summary, we performed biotin-tagging immunoprecipitation with high-throughput sequencing (Bio-ChIP-seq) experiments on Pho2, Pho4, Cbf1, and mock samples in both high-and no-Pi conditions (60 min after Pi starvation), and on Pho4 in a *pho2Δ* strain grown in no-Pi conditions, and in *pho80Δ* and *cbf1Δ pho80Δ* strains grown in high-Pi conditions. Sequencing libraries were prepared for both ChIP DNA and input DNA (from the supernatant of total cell lysate) according to the Illumina protocol. Libraries with size between 200 and 300 bp were selected for PCR amplification and sequenced with an Illumina Genome analyzer II. Thirty-six base sequence tags were aligned to the *S. cerevisiae* genome with ELAND. On average, 2.9 and 12.5 million uniquely aligned reads were obtained for ChIP and input samples, respectively. Bio-ChIP-seq results of all Pho4 consensus binding sites are presented in Table S2.

In Vivo Nucleosome Mapping

Libraries from mononucleosomal DNA were sequenced from both ends with an Illumina genome analyzer II; details are described in the Supplemental Experimental Procedures. We assumed that the center of each sequenced DNA fragment is the nucleosome dyad and extended 73 bp on both sides to generate mononucleosome coverage (Figures S1C–S1E). We obtained nucleosome maps for the wild-type strain in high- and no-Pi conditions (40 min after Pi starvation) and for the *cbf1Δ* strain in high-Pi conditions. For each sample, 8–10 million uniquely aligned nucleosomal DNA sequencing reads were obtained. Nucleosome occupancy of all Pho4 consensus binding sites is presented in Table S2.

Microarray and Data Processing

Details of yeast cell collection, RNA isolation, complementary DNA (cDNA) synthesis and labeling, and microarray hybridization are described in the Supplemental Experimental Procedures. Whole-genome expression profiling was performed with competitive hybridization to Agilent 8 × 15K *S. cerevisiae* two-color expression microarrays (G2509F, AMADID #019838). Microarrays were scanned with an Axon 4000B scanner, and gene features were extracted with GenePix 5.1 software. Lowess and quantile normalization

Supplementary Figure 4 (continued).

were performed with the MATLAB bioinformatics toolbox before further analysis.

Wild-type (WT) no-Pi versus WT high-Pi, *cbf1Δ* high-Pi versus WT high-Pi, *cbf1Δ pho4Δ* high-Pi versus WT high-Pi, and *tye7Δ rtg3Δ* high-Pi versus WT high-Pi microarrays were performed with dye swaps to eliminate dye labeling bias and were analyzed in four biological replicates (Churchill, 2002; Yang and Speed, 2002). *pho80Δ* versus *pho80Δ pho4Δ* and *pho80Δ cbf1Δ* versus *pho80Δ cbf1Δ pho4Δ* were performed in high-Pi conditions and analyzed in two biological replicates. Mutant cycle analysis was constructed with a cyclic comparison so that the expression components could be directly inferred and dye labeling bias would be cancelled in the analysis (Churchill, 2002; Quackenbush, 2002; Yang and Speed, 2002) (Figure S4). The mutant cycle was repeated with three biological replicates.

Mutant Cycle Analysis, also known as Epistasis Expression Analysis

Details of the design and analysis of the mutant cycle (Figure S4) are described in the Supplemental Experimental Procedures. Eighty genes were selected as Pi starvation-responsive genes by direct comparison of the expression of the wild-type in no- and high-Pi conditions (activated more than 1.8-fold and significantly induced with a null hypothesis of < 1.5 fold, $p \leq 0.01$). For each of those genes, linear regression was performed and genes with significant Pho2, Pho4, or Co components (significantly induced with a null hypothesis of < 1.4 fold, $p \leq 0.05$) (Capaldi et al., 2008) are shown in Figure 4A. None of the genes induced after Pi starvation contains a significant Pho2 only component, indicating that Pho4 is the primary activator and Pho2 functions only as a cooperating factor (Komeili and O'Shea, 1999; Springer et al., 2003). Full mutant cycle analysis results are summarized in Table S3.

Determining Whether Binding Sites Are Accessible or Inaccessible

We calculate the average nucleosome occupancy of a 20 bp window centered on all consensus "CACGTG" sites to determine local nucleosome occupancy. Approximately 80% of the *S. cerevisiae* genome is estimated to be covered with nucleosomes (Lee et al., 2007); we thus say that a transcription factor binding site is in the accessible state if the average nucleosome occupancy on the site is in the lower quartile of the average genome nucleosome occupancy. Symmetrically, we say that the site is inaccessible if the average nucleosome occupancy is in the upper quartile of the average genome nucleosome occupancy. The average genome nucleosome occupancy distribution was determined by random sampling of the 20 bp window average for 10,000 genome locations and repetition of this ten times. The average of the upper and lower quartile cutoff for ten-time sampling of the average genome nucleosome occupancy was used to define the accessible and inaccessible states of binding sites. Varying the percentile threshold over a wide range (15%–30%) does not affect the outcome (Figure S2A).

Determining Binding at Consensus CACGTG Sites

Methods to determine the binding at Pho4 consensus binding sites are described in the Supplemental Experimental Procedures.

Pho2 and Pho4 Binding Cooperativity

The increase in Pho2 and Pho4 binding after Pi starvation (recruitment) is calculated with the following equation: taking Pho4 as an example, $\text{Enrichment}_{\text{No4-High}} = (\text{Pho4}_{\text{ChIP No}} / \text{Pho4}_{\text{Input No}}) / (\text{Pho4}_{\text{ChIP High}} / \text{Pho4}_{\text{Input High}})$. To estimate a threshold to define transcription factor recruitment, we calculated the mean and standard deviation of Pho2 and Pho4 recruitment at all consensus "CACGTG" sites, excluding the sites at regulatory regions of Pho4 regulated genes (functional binding sites). We used mean + 2 standard deviations (SDs) as the threshold to identify the sites showing the most significant recruitment (Figure 4B). 23 of 28 sites showing recruitment of both Pho2 and Pho4 are regulated by both factors ($p = 8.1 \times 10^{-16}$, Fisher's exact test), and none of the Pho4-only recruited sites is regulated by Pho4 ($p = 0.003$, Fisher's exact test).

Pho4 Binding at Nonconsensus Binding Sites

Information about Pho4 binding at its nonconsensus binding sites is described in the Supplemental Experimental Procedures.

Fluorescence Microscopy

The procedures for fluorescence microscopy are described in the Supplemental Experimental Procedures.

Prediction of Pho4 Binding and Function

To predict Pho4 binding at its consensus binding motifs, we integrate information about Pho4 DNA binding preference, local nucleosome occupancy, and competition from Cbf1 into an equilibrium model (Granek and Clarke, 2005). At the equilibrium state, the probability of Pho4 not binding to an 8-mer sequence at position i can be expressed as

$$P_{unbound,\,i} = \frac{W_{Nuc,i} + W_{Cbf1,i} + 1}{W_{Nuc,i} + W_{Cbf1,i} + K_{a,\,i,\,adj} * [\text{Pho4}] + 1},$$

where $W_{Nuc,\,i}$, $W_{Cbf1,\,i}$, $K_{a,\,i,\,adj}$ and [Pho4] represent occlusion from nucleosomes, competition from Cbf1, equilibrium association constant of Pho4 with the 8-mer sequence, and the nuclear Pho4 concentration.

If there are n potential Pho4 binding sites near position x on the genome, the binding probability at x is finally calculated as

$$P_{bound,\,x} = 1 - \Pi\, P_{unbound,\,i} (i = 1, 2, 3 \ldots n).$$

Derivation of equations and details about the calculation are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Agilent custom 8 × 15K S. cerevisiae two-color whole genome expression microarray data are deposited in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE23580. All sequencing data are deposited in the GEO database under accession number GSE29506.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be found with this article online at doi:10.1016/j.molcel.2011.05.025.

ACKNOWLEDGMENTS

We thank C. Daly and J. Zhang for help with Illumina sequencing; X. Zhang for assistance with microscopy; former and present O'Shea lab members for discussion and commentary; and V. Denic, N. Friedman, A. Regev, and B. Stern for critical reading of the manuscript. This work was supported by National Institutes of Health grant R01 GM051377 and the Howard Hughes Medical Institute.

Received: October 7, 2010
Revised: January 8, 2011
Accepted: May 18, 2011
Published: June 23, 2011

REFERENCES

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol. Cell *32*, 878–887.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Barbarić, S., Münsterkötter, M., Svaren, J., and Hörz, W. (1996). The homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast PHO5 promoter. Nucleic Acids Res. *24*, 4479–4486.

Supplementary Figure 4 (continued).

Barbaric, S., Münsterkötter, M., Goding, C., and Hörz, W. (1998). Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol. Cell. Biol. *18*, 2629–2639.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Capaldi, A.P., Kaplan, T., Liu, Y., Habib, N., Regev, A., Friedman, N., and O'Shea, E.K. (2008). Structure and function of a transcriptional network activated by the MAPK Hog1. Nat. Genet. *40*, 1300–1306.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. Nat. Genet. Suppl. *32*, 490–495.

Farnham, P.J. (2009). Insights from genomic profiling of transcription factors. Nat. Rev. Genet. *10*, 605–616.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Granek, J.A., and Clarke, N.D. (2005). Explicit equilibrium modeling of transcription-factor binding and gene regulation. Genome Biol. *6*, R87.

Graumann, J., Dunipace, L.A., Seol, J.H., McDonald, W.H., Yates, J.R., 3rd, Wold, B.J., and Deshaies, R.J. (2004). Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast. Mol. Cell. Proteomics *3*, 226–237.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99–104.

Hochheimer, A., and Tjian, R. (2003). Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. *17*, 1309–1320.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Jimenez-Sanchez, G., Childs, B., and Valle, D. (2001). Human disease genes. Nature *409*, 853–855.

Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E., et al. (2010). Variation in transcription factor binding among humans. Science *328*, 232–235.

Kent, N.A., Tsang, J.S., Crowther, D.J., and Mellor, J. (1994). Chromatin structure modulation in Saccharomyces cerevisiae by centromere and promoter factor 1. Mol. Cell. Biol. *14*, 5229–5241.

Kent, N.A., Eibert, S.M., and Mellor, J. (2004). Cbf1p is required for chromatin remodeling at promoter-proximal CACGTG motifs in yeast. J. Biol. Chem. *279*, 27116–27123.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259–272.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. BMC Mol. Biol. *10*, 6.

Komeili, A., and O'Shea, E.K. (1999). Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. Science *284*, 977–980.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285–294.

Lam, F.H., Steger, D.J., and O'Shea, E.K. (2008). Chromatin decouples promoter threshold from dynamic range. Nature *453*, 246–250.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Lee, T.A., Jorgensen, P., Bognar, A.L., Peyraud, C., Thomas, D., and Tyers, M. (2010). Dissection of combinatorial control by the Met4 transcriptional complex. Mol. Biol. Cell *21*, 456–469.

Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D., and Lieb, J.D. (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res. *16*, 1517–1528.

Maclsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics *7*, 113.

Maerkl, S.J., and Quake, S.R. (2007). A systems approach to measuring the binding energy landscapes of transcription factors. Science *315*, 233–237.

Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475–487.

O'Neill, E.M., Kaffman, A., Jolly, E.R., and O'Shea, E.K. (1996). Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science *271*, 209–212.

Ogawa, N., DeRisi, J., and Brown, P.O. (2000). New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. Mol. Biol. Cell *11*, 4309–4321.

Pan, Y., Tsai, C.J., Ma, B., and Nussinov, R. (2010). Mechanisms of transcription factor selectivity. Trends Genet. *26*, 75–83.

Pierce, M., Benjamin, K.R., Montano, S.P., Georgiadis, M.M., Winter, E., and Vershon, A.K. (2003). Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. Mol. Cell. Biol. *23*, 4814–4825.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569–577.

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat. Genet. Suppl. *32*, 496–501.

Robinson, K.A., and Lopes, J.M. (2000). SURVEY AND SUMMARY: Saccharomyces cerevisiae basic helix-loop-helix proteins regulate diverse biological processes. Nucleic Acids Res. *28*, 1499–1505.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Schneider, K.R., Smith, R.L., and O'Shea, E.K. (1994). Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. Science *266*, 122–126.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. *2*, 827–837.

Springer, M., Wykoff, D.D., Miller, N., and O'Shea, E.K. (2003). Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. PLoS Biol. *1*, E28.

van Werven, F.J., and Timmers, H.T. (2006). The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation. Nucleic Acids Res. *34*, e33.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Supplementary Figure 4 (continued).

Vogel, K., Hörz, W., and Hinnen, A. (1989). The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. Mol. Cell. Biol. *9*, 2050–2057.

Wasson, T., and Hartemink, A.J. (2009). An ensemble model of competitive multi-factor binding of the genome. Genome Res. *19*, 2101–2112.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., et al. (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J. *29*, 2147–2160.

Yang, Y.H., and Speed, T. (2002). Design issues for cDNA microarray experiments. Nat. Rev. Genet. *3*, 579–588.

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M., and Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. Nature *464*, 1187–1191.

Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. *19*, 556–566.

Supplementary Figure 4 (continued).